

# Applications of classical approximation theory to periodic basis function networks and computational harmonic analysis

Hrushikesh N. Mhaskar · Paul Nevai · Eugene Shvarts

Received: 22 January 2013 / Revised: 13 June 2013 / Accepted: 17 June 2013 /

Published online: 9 July 2013

© The Author(s) 2013. This article is published with open access at SpringerLink.com

**Abstract** In this paper, we describe a novel approach to classical approximation theory of periodic univariate and multivariate functions by trigonometric polynomials. While classical wisdom holds that such approximation is too sensitive to the lack of smoothness of the target functions at isolated points, our constructions show how to overcome this problem. We describe applications to approximation by periodic basis function networks, and indicate further research in the direction of Jacobi expansion and approximation on the Euclidean sphere. While the paper is mainly intended to be

---

Communicated by S.K. Jain.

---

The research of H.N. Mhaskar was supported, in part, by grant DMS-0908037 from the National Science Foundation and grant W911NF-09-1-0465 from the U.S. Army Research Office. The research of P. Nevai was supported by a KAU grant.

---

H. N. Mhaskar

Department of Mathematics, California Institute of Technology, Pasadena, CA 91125, USA  
e-mail: hmhaska@gmail.com

H. N. Mhaskar

Institute of Mathematical Sciences, Claremont Graduate University, Claremont, CA 91711, USA

P. Nevai (✉)

King Abdulaziz University, Jeddah, Saudi Arabia  
e-mail: paul@nevai.us

E. Shvarts

Shvarts Scientific Services, Pacoima, CA 91331, USA

E. Shvarts

Department of Mathematics, University of California at Davis, Davis, CA 95616, USA  
e-mail: eugene.shvarts@gmail.com

a survey of our recent research in these directions, several results are proved for the first time here.

## 1 Introduction

Two of the major developments during the last quarter of a century are the use of radial basis function (RBF) networks in learning theory, and wavelet analysis in computational harmonic analysis. The subject of radial basis function networks is highly studied, with applications in various fields of mathematics, sciences, engineering, biology, learning theory, and so forth. A *MathSciNet* and a *Web of Science* search on January 10 of 2013 for “radial basis function\*” showed 1,262 and 12,697 citations, respectively. A similar search with “wavelet\*” revealed 10,842 and 65,271 citations, respectively. Therefore, we will make no effort even to attempt to survey the entire subject, and, instead, we will focus on certain facts which have impacted our own research in these areas.

Please note that, for the sake of clarity, the notation used in the introduction may not be the same during the rest of the paper.

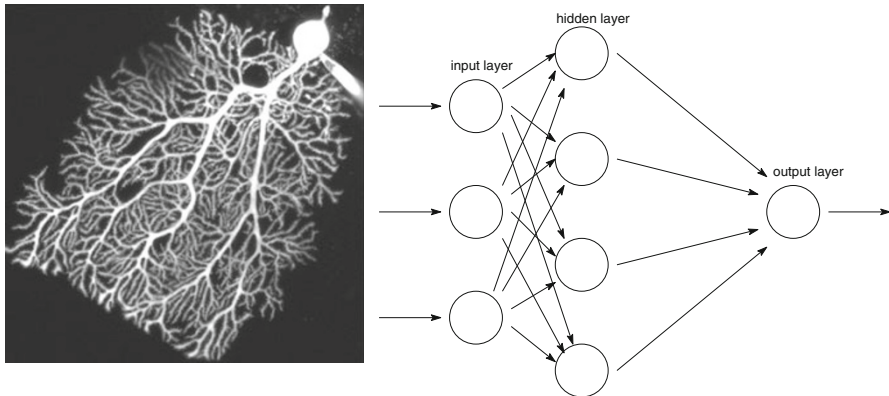
A major theme in learning theory is to discover a functional relationship in given data of the form  $\{(\mathbf{x}_k, y_k)\}_{k=1}^M$  where  $\mathbf{x}_k$ ’s are vectors in a Euclidean space  $\mathbb{R}^q$  for some  $q \in \mathbb{N}$ , and  $y_k$ ’s are the corresponding function values, usually corrupted with noise. The problem is to find a *target function* (or *model*)  $f : \mathbb{R}^q \rightarrow \mathbb{R}$  so that  $f(\mathbf{x}_k) = y_k$ ,  $k = 1, 2, \dots, M$  or at least,  $f(\mathbf{x}_k) \approx y_k$ ,  $k = 1, 2, \dots, M$ . The first requirement is usually imposed either when the data is insufficient, or when it is important to reproduce the data exactly in the model. For example, in image registration applications, one needs to preserve some “landmarks”. On the other hand, if the data is plentiful, but noisy, one does not wish to reproduce it exactly, but wishes to “fit” a smooth model to the data.

In either case, the problem is clearly ill-posed—there are infinitely many models  $f$  which meet the requirements. Therefore, the usual method to obtain the model is first to choose a class  $\mathcal{M}$  of desired models, and find the desired model  $f$  so as to minimize a *regularization functional* such as

$$\sum_{k=1}^M (g(\mathbf{x}_k) - y_k)^2 + \delta \|\mathcal{L}g\| \quad (1.1)$$

over all  $g \in \mathcal{M}$ , where  $\mathcal{L}$  is a *penalty functional* (usually, a differential operator),  $\|\circ\|$  is a suitable norm, and  $\delta$  is the *regularization parameter*. The least square fit is obtained by setting  $\delta = 0$ ; letting  $\delta \rightarrow \infty$  is equivalent to minimizing  $\|\mathcal{L}g\|$  subject to interpolatory conditions. A very classical example of this approach has  $\mathbf{x}_k \in [-1, 1]$ ,  $\mathcal{M}$  is the class of all twice continuously differentiable functions,  $\|\circ\|$  is the  $L^2$  norm on  $\mathbb{R}$ ,  $\mathcal{L}g = g''$ , and we let  $\delta \rightarrow \infty$ . In this case, one recovers the cubic spline interpolant for the data. A technique well-known in image processing is the TV minimization, where  $\mathbf{x}_k \in [0, 1]^2$ ,  $\|\circ\|$  is the  $L^1$  norm, and  $\mathcal{L}g$  is the so-called total variation of  $g$ .

Motivated by the example of spline functions, regularization or smoothing interpolation is used with many other penalty functionals in learning theory. In view of the



**Fig. 1** The image on the left depicts a rat neuron. Neurons fire when their electric potential exceeds a certain threshold value. The term “neural network” arises in conjunction with the similarity found between the architecture of actual neuron cells, and the topology of an artificial neural RBF network, pictured on the right. The input layer receives a vector, and passes it as an argument to each of the small computers represented by circles in the hidden layer. Each of these computers evaluates a term of the form  $G(\mathbf{x} - \mathbf{x}_j)$ , and represents the analogous dependence on electric potential. The output layer takes the linear combination. The image copyrights are held by Testuya Tatsukawa 2010 and Unikom Center 2010–2012, respectively

Golomb–Weinberger principle, the solution to these problems can often be obtained explicitly in the form  $\sum_{k=1}^M a_k \phi(|\circ - \mathbf{x}_k|)$  for a univariate function  $\phi$ , where  $|\circ|$  denotes the usual Euclidean norm [15, 25]. A function of this form is called a *radial basis function (RBF) network* with  $M$  neurons. The function  $\phi$  is called the *activation function* and the  $\mathbf{x}_k$ ’s are the *centers* of the network. More generally, a *translation network* is a function of the form  $\sum_{k=1}^M a_k G(\circ - \mathbf{x}_k)$  where the activation function  $G$  is defined on an appropriate Euclidean space.

The motivation for this terminology stems from an analogy with neurons in the brains, as indicated by Fig. 1.

A question of central interest in machine learning can be formulated as a problem of function approximation. A very classical theorem of Wiener states that the set of all translation networks is dense in  $L^1(\mathbb{R}^q)$  if and only if  $\hat{G}(\mathbf{t}) \neq 0$  for any  $\mathbf{t} \in \mathbb{R}^q$ . Under mild conditions on  $\phi$ , it was proved by Park and Sandberg [70] that the class of all RBF networks

$$\left\{ \sum_{k=1}^M a_k \phi(|\circ - \mathbf{x}_k|) : M \geq 1, \mathbf{x}_k \in \mathbb{R}^q \right\}$$

is dense in the space of continuous real valued functions on  $\mathbb{R}^q$  with the topology of convergence on compact subsets.

In [50], we showed that the set of all translation networks with activation function  $G$  which can be interpreted as a tempered distribution is dense in the same sense if and only if the support of the distributional Fourier transform of  $G$  is a so-called “set of uniqueness” for entire functions of finite exponential type in  $q$  variables.

Perhaps, the most popular way of using translation networks is for interpolating data. An attractive feature of RBF networks is a result by Micchelli [62] that (again under

mild conditions on  $G$ ) the matrix  $[G(\mathbf{x}_j - \mathbf{x}_k)]_{j,k=1}^M$  is always invertible for an arbitrary choice of  $M$  and  $\mathbf{x}_k$ 's, so that interpolation by RBF networks is always possible. It is proved in [25] that an interpolatory network also solves certain regularization problems.

While interpolation by RBF networks as well as the use of extremal problems to obtain a functional relationship underlying a data are very popular methods, there are some disadvantages. First, there is no guarantee that as the data increases, the minimal value of the regularization functional will remain bounded. Second, there are usually no a priori bounds on the accuracy of the resulting approximation. Indeed, the theory of degree of approximation by interpolatory RBF networks is a fairly well established topic in its own right. Finally, there are numerous computational issues, including ill\*-conditioning, lack of convergence, local minima, and so forth.

In the past 20 years or so, the first author, together with collaborators, has explored methods to construct translation network approximations with a priori performance guarantees which avoid each of the pitfalls mentioned above. Quite often, the resulting networks also satisfy the extremal properties up to a constant multiple. This research has demonstrated a very close connection between the classical theory of polynomial approximation, and approximation by translation networks. Applications of the ideas in this research include the theory of probability density estimation [78], pattern classification [3], control theory [33], signal processing [16], numerical simulation of turbulent channel flow [23], and construction of Schauder basis [24, 37, 31].

Some of the highlights of our work are the following:

- The conditions on the activation function required to achieve our approximation bounds are weaker than those required for interpolation.
- Our procedures are given as explicit formulas, requiring neither training in the classical machine learning setting, nor a solution of a system of equations involving a possibly ill-conditioned large matrix. The formulas can be implemented as a matrix–vector multiplication.
- The spaces to which the target function is assumed to belong are the usual smoothness classes, rather than the so-called native spaces for the networks.
- It is easy to adopt a two-scale approach to constructing a network which provides both the optimal approximation bounds, and interpolation at fewer points than there are centers of the network (cf. [7]).

The purpose of this paper is to illustrate the main ideas in our research in a few contexts. In Sect. 2, we introduce the basic ideas in the context of approximation by univariate trigonometric polynomials. The material in this section is extended in the context of multivariate trigonometric polynomials in Sect. 3. A new feature here is the ability to approximate functions based on “scattered data”, that is, evaluations where one does not prescribe the location of the points where the function is to be evaluated. The analogues of these results in the context of periodic basis function networks are given in Sect. 4. Certain extensions of the various parts of this theory are discussed in Sect. 5.

## 2 Approximation of univariate periodic functions

### 2.1 Preliminaries

In this section we describe some basic results regarding approximation of univariate  $2\pi$ -periodic functions by trigonometric polynomials. There are numerous standard references on the subject (for instance, [6, 12, 57, 66, 77]). Our summary in this section is based on [57], where many results are given with elementary proofs for the case of uniform approximation, and [12], where these results are given in the full generality. First, some terminology. We denote by  $\mathbb{T}$  the quotient space of the interval  $[-\pi, \pi]$  where the end points are identified. Geometrically, we think of  $\mathbb{T}$  as the unit complex circle, except that rather than denoting a point on  $\mathbb{T}$  by  $e^{ix}$ , we simplify the notation and denote it by  $x$ . If  $f : \mathbb{T} \rightarrow \mathbb{R}$  is Lebesgue measurable, and  $A \subseteq \mathbb{T}$  is Lebesgue measurable, we write

$$\|f\|_{p,A} \stackrel{\text{def}}{=} \begin{cases} \left\{ \frac{1}{2\pi} \int_A |f(t)|^p dt \right\}^{1/p}, & \text{if } 0 < p < \infty, \\ \text{ess sup}_{t \in A} |f(t)|, & \text{if } p = \infty. \end{cases} \quad (2.1)$$

The set of all Lebesgue measurable functions for which  $\|f\|_{p,A} < \infty$  is denoted by  $L^p(A)$ , with the understanding that functions which are equal almost everywhere on  $A$  are considered equal as members of  $L^p(A)$ . The set of all uniformly continuous, bounded, and  $2\pi$ -periodic functions on  $A$ , equipped with the norm of  $L^\infty(A)$  (which is known as the *uniform* or *supremum* norm in this case) is denoted by  $C^*(A)$ . If  $A = \mathbb{T}$ , we will omit it from the notation. In the sequel, we will assume that  $1 \leq p \leq \infty$ . The *dual exponent notation* is as follows:

$$p' \stackrel{\text{def}}{=} \begin{cases} p/(p-1), & \text{if } 1 < p < \infty, \\ \infty, & \text{if } p = 1, \\ 1, & \text{if } p = \infty. \end{cases}$$

If  $n$  is a non-negative integer, then a *trigonometric polynomial of degree (or order)  $n$*  is defined to be a function of the form

$$x \mapsto \sum_{k=-n}^n c_k e^{ikx}, \quad x \in \mathbb{T},$$

where  $c_k$ 's are complex numbers, known as the coefficients of the polynomial, and  $|c_n| + |c_{-n}| \neq 0$ . The set of all trigonometric polynomials of degree  $< n$  will be denoted by  $\mathbb{H}_n$ . It is convenient to extend this notation to non-integer values of  $n$  by setting  $\mathbb{H}_n \stackrel{\text{def}}{=} \mathbb{H}_{[n]}$  where  $[n]$  is the integer part of  $n$ , and setting  $\mathbb{H}_n = \{0\}$  if  $n \leq 0$ . According to the trigonometric variant of the Weierstrass theorem and its  $L^p$ -versions, the  $L^p$ -closure of  $\cup_{n=1}^\infty \mathbb{H}_n$  is  $L^p$  for  $1 \leq p < \infty$  and  $C^*$  for  $p = \infty$ . In order to avoid making an elaborate distinction between the two cases every time we state a theorem, we will write  $X^p = L^p$  if  $p < \infty$  and  $X^\infty \stackrel{\text{def}}{=} C^*$ . A central theme of approximation

theory in this context is to investigate the properties of the *degree of approximation* of  $f \in X^p$  from  $\mathbb{H}_n$  for all  $n$ . This is defined for  $f \in X^p$  and  $n \geq 0$  by

$$E_{n,p}(f) \stackrel{\text{def}}{=} \inf\{\|f - P\|_p : P \in \mathbb{H}_n\}.$$

The quantity  $E_{n,p}(f)$  measures the minimal error that one **must** expect if one wishes to use an element of  $\mathbb{H}_n$  as a model for  $f$ . Clearly, the function  $n \mapsto E_{n,p}(f)$  is non-increasing, the function  $p \mapsto E_{n,p}(f)$  is non-decreasing, and the function  $f \mapsto E_{n,p}(f)$  is a semi-norm on  $X^p$  if  $p \geq 1$ . In the sequel, we will assume that  $1 \leq p \leq \infty$ .

Two of the main problems in this theory are the following. One of them is how to construct the *best approximation*  $P_p^*(f) \in \mathbb{H}_n$  such that  $\|f - P_p^*(f)\|_p = E_{n,p}(f)$ . The other problem is to give, for  $\gamma > 0$ , a complete characterization of  $f \in X^p$  for which  $E_{n,p}(f) = \mathcal{O}(n^{-\gamma})$ . Except in the case when  $p = 2$ , the operator  $P_p^*$  is non-linear, and it takes a very elaborate optimization technique to compute  $P_p^*(f)$  in general. Thus we adopt the stance that it is computationally worthwhile to construct *near-best approximations* or *sub-optimal solutions*, that is, finding some  $P \in \mathbb{H}_n$  for which  $\|f - P\|_p \leq c_1 E_{n,p}(f)$  for some positive constants  $c, c_1$  independent of  $n$  or  $f$ . We discuss these constructions in this subsection, postponing the discussion of the characterization question to Sect. 2.2.

### 2.1.1 Constant convention

In the remainder of this paper, the symbols  $c, c_1, \dots$ , will denote generic positive constants whose value is independent of the target function  $f$ , and other variable parameters such as  $n$ , but may depend on fixed parameters under discussion, such as the norm  $p$  or the smoothness index  $\gamma$ , and so forth. Their value may be different at different occurrences, even within a single formula. The notation  $A \sim B$  will mean  $c_1 A \leq B \leq c_2 A$ . As usual, the notation  $A = \mathcal{O}(B)$  will mean that  $|A| \leq c|B|$ , where  $c$  is some positive constant whose value may depend on  $f$  which one way or another may appear in the expressions  $A$  and  $B$ .

In the case when  $p = 2$ , an explicit formula for the best approximation polynomial  $P_2^*(f)$  is well known. We define the *Fourier coefficients* of  $f$  by

$$\hat{f}(k) = \frac{1}{2\pi} \int_{\mathbb{T}} f(t) \exp(-ikt) dt, \quad k \in \mathbb{Z},$$

and for  $n \in \mathbb{N}$  we set

$$s_n(f, x) = \sum_{|k| \leq n-1} \hat{f}(k) \exp(ikx), \quad x \in \mathbb{T}.$$

It is well known that  $P_2^*(f) = s_n(f)$ , that is,

$$\|s_n(f) - f\|_2 = \inf\{\|f - P\|_2 : P \in \mathbb{H}_n\}.$$

It is well known [79, Chapter VII, Theorem 6.4] that

$$\|s_n(f) - f\|_p \leq cE_{n,p}(f), \quad 1 < p < \infty,$$

where  $c$  is a constant depending on  $p$ . The value of  $c$  tends to  $\infty$  if  $p \rightarrow 1, \infty$ . There exist integrable functions  $f$  for which the sequence  $\{s_n(f, x)\}$  diverges for almost all  $x$ , and a dense set of functions  $f \in C^*$  for which  $\{s_n(f, 0)\}$  diverges. A very deep theorem in the theory of Fourier series, the Carleson–Hunt theorem [4, 30] states that if  $p > 1$  and  $f \in L^p$  then the sequence  $\{s_n(f, x)\}$  converges for almost all  $x$  to  $f(x)$ .

A ground-breaking theorem in this direction states that if

$$\sigma_n(f, x) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{m=1}^n s_m(f, x), \quad n \in \mathbb{N} \text{ and } x \in \mathbb{T},$$

then for all  $p$  with  $1 \leq p \leq \infty$  and  $f \in X^p$ ,

$$\lim_{n \rightarrow \infty} \|f - \sigma_n(f)\|_p = 0.$$

In the case when  $p = \infty$ , this theorem was proved by Fejér in [18] and it appears in almost every textbook on approximation theory, see e.g., [12, Chapter 1, Corollary 2.2] or [57, Chapter 1, Section 1.1]. In the case of  $1 \leq p < \infty$  and, in fact, in greater generality, it appears in [79, Chapter IV, Theorem 5.14]. The rate of convergence

$$\|f - \sigma_n(f)\|_p \leq \frac{c}{n} \sum_{k=1}^n E_{k,p}(f),$$

was given in 1961 by Stechkin (aka Stečkin), see [74]. In order to get a near-best approximation, we define

$$v_n(f, x) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{m=n+1}^{2n} s_m(f, x), \quad n \in \mathbb{N} \text{ and } x \in \mathbb{T},$$

where  $v$  is in honor of C. de La Vallée Poussin. It is easy to check that

$$v_n(f) = 2\sigma_{2n}(f) - \sigma_n(f).$$

It can be deduced from here ([12, Chapter 9, Theorem 3.1]) that for all  $p$  such that  $1 \leq p \leq \infty$  and for all  $f \in L^p$  one has

$$E_{2n,p}(f) \leq \|f - v_n(f)\|_p \leq 4E_{n,p}(f).$$

To understand the difference between the behavior of  $\sigma_n(f)$  and  $v_n(f)$ , we point out an alternative expression for these. Namely,

$$\begin{aligned}\sigma_n(f, x) &= \sum_{|k| < n} \left(1 - \frac{|k|}{n}\right) \hat{f}(k) \exp(ikx), \\ v_n(f, x) &= \sum_{|k| \leq n} \hat{f}(k) \exp(ikx) + \sum_{n+1 \leq |k| < 2n} \left(2 - \frac{|k|}{n}\right) \hat{f}(k) \exp(ikx).\end{aligned}\quad (2.2)$$

Thus, if  $P \in \mathbb{H}_n$  and  $P$  is not a constant, then  $\sigma_n(P) \neq P$  but  $v_n(P) \equiv P$ . We observe further that  $v_n(f)$  can be written in the form

$$v_n(f, x) = \sum_{|k| < 2n} h\left(\frac{k}{2n}\right) \hat{f}(k) \exp(ikx),$$

where

$$h(t) = \begin{cases} 1, & \text{if } |t| \leq 1/2, \\ 2 - 2t, & \text{if } 1/2 < |t| < 1, \\ 0, & \text{if } |t| \geq 1. \end{cases}$$

Motivated by this observation, we make the following definition.

**Definition 2.1** Let  $h : \mathbb{R} \rightarrow [0, 1]$  be compactly supported. The **summability operator** (corresponding to the **filter**  $h$ ) is defined for  $f \in L^1$  by

$$\sigma_n(h, f, x) \stackrel{\text{def}}{=} \sum_{k \in \mathbb{Z}} h\left(\frac{k}{n}\right) \hat{f}(k) \exp(ikx), \quad n \in \mathbb{N} \text{ \& } x \in \mathbb{T}. \quad (2.3)$$

The **summability kernel** corresponding to  $h$  is defined by

$$\Phi_n(h, t) \stackrel{\text{def}}{=} \sum_{k \in \mathbb{Z}} h\left(\frac{k}{n}\right) \exp(ikt), \quad t \in \mathbb{T}. \quad (2.4)$$

The function  $h$  is called a **low pass filter** if  $h$  is an even function, non-increasing on  $[0, \infty)$ ,  $h(t) = 1$  if  $|t| \leq 1/2$ , and  $h(t) = 0$  if  $|t| \geq 1$ .

We note that for  $n \in \mathbb{N}$  and  $f \in L^1$ ,

$$\begin{aligned}\sigma_n(h, f, x) &= \frac{1}{2\pi} \int_{\mathbb{T}} f(t) \Phi_n(h, x - t) dt \\ &= \frac{1}{2\pi} \int_{\mathbb{T}} f(x - t) \Phi_n(h, t) dt, \quad f \in L^1, \quad x \in \mathbb{T}.\end{aligned}\quad (2.5)$$



Even though the sum in (2.3) is written as an infinite sum, it is in fact a finite sum since  $h(|k|/n) = 0$  if  $|k|$  is sufficiently large; if  $h$  is a low pass filter, then  $k \geq n$  is sufficient. Therefore, for such  $h$ ,  $\sigma_n(h, f) \in \mathbb{H}_n$ . By writing the sum as an infinite sum, we avoid the need to restrict ourselves in the definition to the case when  $n$  is an integer. If  $h$  is supported on  $[-A, A]$ , then  $\Phi_{1/A}(h, t) \equiv 1$ , and we redefine  $\Phi_n(h, t) \equiv 0$  if  $n < 1/A$ .

We summarize some important properties of the summability operator. In what follows, the constants may depend on the function  $h$ .

**Theorem 2.1** (a) *If  $S \geq 2$  is an integer, and  $h$  is an  $S$ -times continuously differentiable function, then*

$$|\Phi_n(h, t)| \leq cn \min \left( 1, (n|t|)^{-S} \right), \quad t \in \mathbb{T} \text{ \& } n \geq 1. \quad (2.6)$$

(b) *Let  $h$  be a twice continuously differentiable, even function,  $1 \leq p \leq \infty$ . Then*

$$\|\sigma_n(h, f)\|_p \leq c\|f\|_p, \quad f \in L^p \text{ \& } n \in \mathbb{N}. \quad (2.7)$$

(c) *Let  $h$  be a twice continuously differentiable low pass filter,  $n \in \mathbb{N}$ . Then  $\sigma_n(h, P) \equiv P$  for all  $P \in \mathbb{H}_{n/2}$ . In addition, for  $1 \leq p \leq \infty$  and  $f \in L^p$ ,*

$$E_{n,p}(f) \leq \|f - \sigma_n(h, f)\|_p \leq cE_{n/2,p}(f). \quad (2.8)$$

Since the localization estimate (2.6) and its analogues in various contexts play an important role in our theory, we pause in our discussion to illustrate it by an example. We consider two low pass filters,  $h_3$  and  $h_\infty$  defined on  $(1/2, 1)$  by

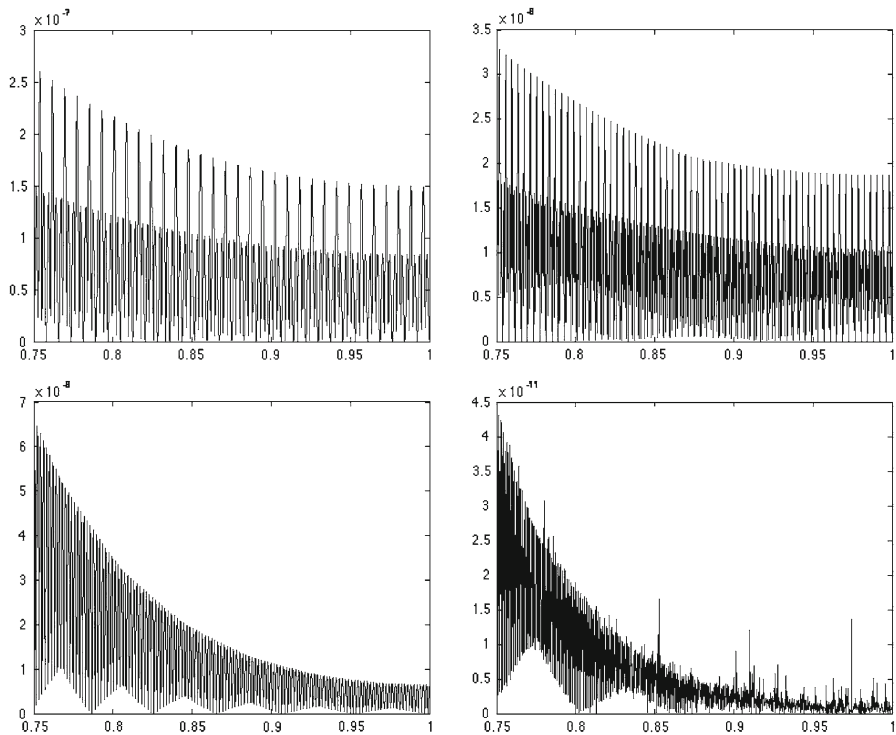
$$h_3(t) = (1-t)^3 \left( 8 + 48(t-1/2) + 192(t-1/2)^2 \right),$$

and

$$h_\infty(t) = \exp \left( -\frac{\exp(2/(1-2t))}{1-t} \right).$$

Of course, both functions are equal to 1 on  $[0, 1/2]$  and 0 on  $[1, \infty)$ . The function  $h_3$  is twice continuously differentiable on  $[0, \infty)$  whereas  $h_\infty$  is infinitely many times differentiable. In Fig. 2, we show the graphs of  $|\Phi_n(h_3, x)|$  and  $|\Phi_n(h_\infty, x)|$  for  $n = 512$  and  $n = 1024$  on the interval  $[\pi/3, \pi]$ . It is clear from the figure that for both the values of  $n$ , the graph corresponding to  $h_\infty$  is an order of magnitude smaller than that for  $h_3$ , and that the graph corresponding to  $h_\infty$  decreases much faster as  $n$  (and/or  $|x|$ ) increases.

*Proof of Theorem 2.1* Part (a) can be proved using the Poisson summation formula; see, e.g., [75, Chapter VII, Theorem 2.4 and Corollary 2.6]; a recent proof is given in



**Fig. 2** Clockwise from the top left, the graphs of  $|\Phi_{512}(h_3, x)|$ ,  $|\Phi_{1024}(h_3, x)|$ ,  $|\Phi_{1024}(h_\infty, x)|$ , and  $|\Phi_{512}(h_\infty, x)|$ . The numbers on the  $x$  axis are in multiples of  $\pi$ . The maximum absolute values of these are  $2.6002e - 07$ ,  $3.2776e - 08$ ,  $4.4293e - 11$ , and  $6.6296e - 08$  respectively

[22]. If  $h$  is twice continuously differentiable, we use (2.6) with  $S = 2$  to obtain

$$\begin{aligned} \int_{\mathbb{T}} |\Phi_n(h, t)| dt &= \int_{|t| \leq 1/n} |\Phi_n(h, t)| dt + \int_{t \in \mathbb{T}, |t| > 1/n} |\Phi_n(h, t)| dt \\ &\leq cn(2/n) + 2cn^{-1} \int_{1/n}^{\pi} t^{-2} dt \leq 2c + 2cn^{-1} \int_{1/n}^{\infty} t^{-2} dt = 4c. \end{aligned}$$

The estimate (2.7) is now deduced using the convolution identity (2.5) and Young's inequality (cf. [79, Chapter II, Theorem 1.15]). This proves part (b). Since  $h$  is a low pass filter,  $h(|k|/n) = 1$  if  $|k| < n/2$ . If  $P \in \mathbb{H}_{n/2}$  then  $\hat{P}(k) = 0$  if  $|k| \geq n/2$ . Hence,

$$\sigma_n(h, P, x) = \sum_{|k| < n/2} h\left(\frac{|k|}{n}\right) \hat{P}(k) \exp(ikx) = P(x), \quad x \in \mathbb{T}.$$

This proves the first assertion in part (c). Let  $f \in L^p$ . The first inequality in (2.8) follows from the fact that  $\sigma_n(h, f) \in \mathbb{H}_n$ . If  $P \in \mathbb{H}_{n/2}$  is arbitrary, then

$$\begin{aligned}\|f - \sigma_n(h, f)\|_p &= \|f - P - \sigma_n(h, f - P)\|_p \leq \|f - P\|_p + \|\sigma_n(h, f - P)\|_p \\ &\leq c_1 \|f - P\|_p.\end{aligned}$$

The second inequality in (2.8) follows by taking the infimum over  $P \in \mathbb{H}_{n/2}$ .  $\square$

## 2.2 Degree of approximation

If  $\epsilon_n \downarrow 0$  as  $n \rightarrow \infty$ , then it is easy to construct  $f \in C^*$  for which  $E_{n,\infty}(f) \leq \epsilon_n$ . We set  $\delta_k = \epsilon_k - \epsilon_{k+1}$  if  $k \geq 0$ , and observe that  $\sum \delta_k$  is a convergent series with non-negative terms. Since  $|\cos(kx)| \leq 1$  for all  $x \in \mathbb{T}$ , the series

$$f(x) = \sum_{k=0}^{\infty} \delta_k \cos(kx)$$

converges uniformly and absolutely to  $f \in C^*$ . Clearly,

$$E_{n,\infty}(f) \leq \sum_{k=n}^{\infty} \delta_k = \epsilon_n.$$

As a matter of fact, as proved by Bernstein (cf. [66, Chapter V, Section 5, p. 109]), one could even have  $E_{n,\infty}(f) = \epsilon_n$ . A central question in approximation theory is to determine the constructive properties (smoothness) of  $f$  which are equivalent to a given rate of decrease of the sequence  $\{E_{n,p}(f)\}$ . In particular, we will examine the very classical case where  $E_{n,p}(f) = \mathcal{O}(n^{-\gamma})$  for some  $\gamma > 0$ .

The fundamental inequalities in this theory are given in the following theorem (for the uniform norm and elementary proofs, see [57, Chapter III, Section 3.1, Theorem 2, Chapter III, Section 2.2, Corollary 3] and in greater generality [12, Chapter 4, Theorem 2.4, Chapter 7, Section 4]). The best constant in the Bernstein inequality, see (2.9) below, is well-known. The idea of using summability operators to prove Bernstein-type inequalities goes back to 1914 when Riesz realized that Landau's trick of applying Fejér's ideas works for Bernstein-type inequalities. Riesz even managed to find the best constant with this seemingly inefficient approach; for the full story see [69]. While Riesz used an explicit expression for the kernel, we will use only the filter which gives rise to the summability kernel. We do not get the best constants, but the method can be generalized considerably to the settings when such an explicit expression for the kernel is not available, see, e.g., [39]. The best constants in (2.10) below are known [12, Chapter 7, Theorem 4.3], [57, Chapter III, Section 2.2, Theorem 4] as well.

**Theorem 2.2** *Let  $1 \leq p \leq \infty$ ,  $n \in \mathbb{N}$ , and let  $r \in \mathbb{N}$ .*

(a) *If  $P \in \mathbb{H}_n$  then the Bernstein-type inequality*

$$\|P^{(r)}\|_p \leq cn^r \|P\|_p \tag{2.9}$$

*holds.*

(b) If  $f \in X^p$  and  $f^{(r)} \in X^p$ , then the Favard-type inequality

$$E_{n,p}(f) \leq \frac{c}{n^r} \|f^{(r)}\|_p. \quad (2.10)$$

holds.

*Proof* Let us fix an infinitely differentiable low pass filter  $h$ . Clearly, the function  $h_1(t) \stackrel{\text{def}}{=} t^r h(t)$ ,  $t \in \mathbb{R}$ , is compactly supported and it is infinitely differentiable. To prove part (a), let  $P \in \mathbb{H}_n$ . Then, in view of Theorem 2.1(c), we have

$$\begin{aligned} P^{(r)}(x) &= \sum_{k \in \mathbb{Z}} h\left(\frac{k}{2n}\right) \widehat{P^{(r)}}(k) \exp(ikx) = \sum_{k \in \mathbb{Z}} h\left(\frac{k}{2n}\right) (ik)^r \hat{P}(k) \exp(ikx) \\ &= (2in)^r \sum_{k \in \mathbb{Z}} h_1\left(\frac{k}{2n}\right) \hat{P}(k) \exp(ikx) = (2in)^r \sigma_{2n}(h_1, P, x). \end{aligned}$$

The estimate (2.9) now follows immediately from (2.7). This proves part (a).

Next, we prove part (b). For  $f \in X^p$ , we define

$$P_j \stackrel{\text{def}}{=} \begin{cases} \sigma_1(h, f), & \text{if } j = 0, \\ \sigma_{2^j}(h, f) - \sigma_{2^{j-1}}(h, f), & \text{if } j \in \mathbb{N}. \end{cases} \quad (2.11)$$

Then for all integer  $m \geq 2$ ,

$$\sum_{j=0}^m P_j = \sigma_{2^m}(h, f), \quad (2.12)$$

and so, since  $f \in X^p$ , (2.8) implies that

$$f = \sum_{j=0}^{\infty} P_j, \quad (2.13)$$

with the series converging in the sense of  $X^p$ .

Next, let  $g(t) = h(t) - h(2t)$ , and  $g_1(t) = g(t)t^{-r}$ . Then  $g(t) = 0$  if  $|t| \leq 1/4$ , and, hence,  $g_1$  is infinitely differentiable. For  $j \in \mathbb{N}$  and  $x \in \mathbb{T}$ , we have

$$\begin{aligned} P_j(x) &= \sum_{k \in \mathbb{Z}} \left( h\left(\frac{|k|}{2^j}\right) - h\left(\frac{|k|}{2^{j-1}}\right) \right) \hat{f}(k) \exp(ikx) \\ &= \sum_{k \in \mathbb{Z}} g\left(\frac{|k|}{2^j}\right) \hat{f}(k) \exp(ikx) = \sigma_j(g, f, x). \end{aligned}$$

Let  $j \geq 2$ . Taking into consideration that  $g(t) = 0$  if  $|t| \leq 1/4$ , we can deduce that

$$\begin{aligned} P_j(x) &= \sigma_{2j}(g, f, x) = \sum_{k \in \mathbb{Z}} g\left(\frac{k}{2^j}\right) \hat{f}(k) \exp(ikx) \\ &= \sum_{k \in \mathbb{Z}} g\left(\frac{k}{2^j}\right) \frac{1}{(ik)^r} \widehat{f^{(r)}}(k) \exp(ikx) \\ &= \frac{1}{(i2^j)^r} \sum_{k \in \mathbb{Z}} g_1\left(\frac{k}{2^j}\right) \widehat{f^{(r)}}(k) \exp(ikx) \\ &= \frac{1}{(i2^j)^r} \sigma_{2j}(g_1, f^{(r)}, x). \end{aligned}$$

Therefore, (2.7) shows that

$$\|P_j\|_p \leq c 2^{-jr} \|f^{(r)}\|_p.$$

Using (2.12) and (2.13), we conclude that

$$E_{2^m, p}(f) \leq \|f - \sigma_{2^m}(h, f)\|_p \leq \sum_{j=m+1}^{\infty} \|P_j\|_p \leq c \|f^{(r)}\|_p \sum_{j=m+1}^{\infty} 2^{-jr} = c_1 2^{-mr} \|f^{(r)}\|_p.$$

If  $n \geq 4$ , we find  $m \geq 2$  such that  $2^m \leq n \leq 2^{m+1}$ . Then the above estimate shows (2.10) in the case when  $n \geq 4$ . The estimate is trivial if  $n = 1, 2, 3$ .  $\square$

We pause in our discussion to introduce the notions of widths in approximation theory, in contradistinction with the characterization theorem (Theorem 2.3 below), which is our main interest in this section.

Let  $X$  be a normed linear space,  $n \geq 1$  be an integer,  $K \subset X$  be a compact set, and  $Y \subset X$  be a closed set. We define the distance

$$\text{dist}(K, Y) \stackrel{\text{def}}{=} \sup_{f \in K} \inf_{P \in Y} \|f - P\|_X. \quad (2.14)$$

In this paper, we will refer to bounds on  $\text{dist}(K, Y)$  as *distance bounds*.

The notion of widths in approximation theory gives in some sense minimal distance of  $K$  from various different choices of  $Y$ . Depending upon the nature of  $Y$ , there are different variants of widths [36, 72]. We will describe two of these.

For integer  $n \geq 1$ , the *Kolmogorov  $n$ -width* of  $K$  [32] is defined by

$$d_{n, \text{kol}}(K, X) \stackrel{\text{def}}{=} \inf_{X_n} \text{dist}(K, X_n), \quad (2.15)$$

where the infimum is taken over all linear subspaces  $X_n$  of  $X$ , with the dimension of  $X_n$  being  $\leq n$ . More generally, any process of approximation of elements of  $K$  based on  $n$  parameters can be formalized as the composition of two functions: the

first,  $\mathcal{M} : K \rightarrow \mathbb{R}^n$ , represents the selection of the parameters, while the second,  $\mathcal{R} : \mathbb{R}^n \rightarrow X$ , represents the reconstruction. Their composition  $f \mapsto \mathcal{R}(\mathcal{M}(f))$  is the desired approximation of  $f \in K$ . The nonlinear  $n$ -width of  $K$  with respect to  $X$  is defined by DeVore et al. [11] and independently by Mathé [41] as

$$d_n(K, X) \stackrel{\text{def}}{=} \inf_{f \in K} \sup \|f - \mathcal{R}(\mathcal{M}(f))\|_X, \quad (2.16)$$

where the infimum is taken over all  $\mathcal{R} : \mathbb{R}^n \rightarrow X$ , and all *continuous*  $\mathcal{M} : K \rightarrow \mathbb{R}^n$ . Here, the reconstruction operation can be arbitrary. The nonlinear  $n$ -width gives the theoretically minimum “worst-case error” in approximating elements of  $K$ , subject only to the prior knowledge that they are elements of  $K$ , using  $n$  parameters selected in a stable manner.

It is not immediately clear that the choice of parameters involved in approximation from an arbitrary finite dimensional space is always continuous. Therefore, it is not obvious that the nonlinear  $n$ -width is a sharpening of the Kolmogorov  $n$ -width. However, in view of [11, Corollary 2.2], it follows that

$$d_n(K, X) \leq d_{n, \text{kol}}(K, X).$$

In many applications,  $K$  is defined in terms of a semi-norm  $\|\cdot\|_K$  on a subset of  $X$ :

$$K = \{f \in X : \|f\|_K \leq 1\}.$$

As a consequence of [11, Theorem 3.1], it follows that if there exists a linear subspace  $Y \subset X$  with dimension  $n + 1$  such that a Bernstein inequality of the form

$$\|g\|_K \leq b_n(K) \|g\|_X, \quad g \in Y, \quad (2.17)$$

holds, then with an absolute constant  $c$ ,

$$d_n(K, X) \geq c b_n(K)^{-1}. \quad (2.18)$$

For example, let  $X = X^p$ ,  $Y = \mathbb{H}_n$ . Then  $\text{dist}(K, \mathbb{H}_n)$  is the “worst-case error” in approximating elements of  $K$  by trigonometric polynomials of degree  $< n$ . When

$$K \stackrel{\text{def}}{=} \{f \in X^p : \|f^{(r)}\|_p \leq 1\}, \quad (2.19)$$

the Favard inequality (2.10) shows the upper distance bound

$$\text{dist}(K, \mathbb{H}_n) \leq c n^{-r}.$$

Therefore, with  $n = 2m - 1$ , we see that  $d_n(K) \leq d_{n, \text{kol}}(K) \leq c n^{-r}$ , while the Bernstein inequality (2.9) and (2.18) together show that  $d_n(K) \geq c n^{-r}$ . Moreover, Theorem 2.1 then leads us to conclude that using the Fourier coefficients for the

parameter selection, followed by the operator  $\sigma_n$  is, up to a constant factor, an optimal reconstruction algorithm for approximation of functions in  $K$ .

The estimates on the  $n$ -widths or the lower distance bounds imply that the degree of approximation or method of approximation cannot be improved, in the sense that there is always some “bad function” in the class for which a lower bound is attained. They do not address the question of whether *individual functions* can be approximated better than what the degree of approximation theorem predicts, based on the a priori information known about the target function. For example, one could conceivably use some clever ideas appropriate for the target function, that can yield a better performance than the theoretically assumed a priori information. This is the question of *characterization of smoothness classes*. This is our main interest in this section, and we now resume this discussion.

It is perhaps clear that the number of derivatives is not a sufficiently sophisticated indication to characterize functions for which  $E_{n,p}(f) = \mathcal{O}(n^{-\gamma})$ ; e.g., when  $\gamma$  is not an integer. If  $\gamma$  is an integer and  $f^{(\gamma)} \in X^p$ , then  $E_{n,p}(f) = \mathcal{O}(n^{-\gamma})$  as proved in (2.10). The converse is not true. For example, if  $f(x) = |\cos x|$ , then it is easy to compute that

$$f(x) = \frac{2}{\pi} + \frac{4}{\pi} \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\cos(2kx)}{4k^2 - 1}$$

where the series converges uniformly. Therefore,

$$E_{2n-1,\infty}(f) \leq \|f - s_{2n}(f)\|_{\infty} \leq \frac{4}{\pi} \sum_{k=n}^{\infty} \frac{1}{4k^2 - 1} \leq c/n.$$

Since  $f$  is not differentiable at  $\pm\pi/2$ , the converse of the Favard estimate is not true.

The correct device is a regularization functional that is known in this context as a  $K$ -functional.

**Definition 2.2** If  $r \in \mathbb{N}$ ,  $1 \leq p \leq \infty$ , and  $f \in X^p$ , then the  $K$ -functional is defined by

$$K_{r,p}(f, \delta) \stackrel{\text{def}}{=} \inf\{\|f - g\|_p + \delta^r \|g^{(r)}\|_p\}, \quad (2.20)$$

where the infimum is taken over all  $g$  for which  $g^{(r)} \in X^p$ . If  $\gamma > 0$  and  $r > \gamma$  is an integer, then we define

$$\|f\|_{\gamma,p} \stackrel{\text{def}}{=} \sup_{0 < \delta < 1/2} \frac{K_{r,p}(f, \delta)}{\delta^{\gamma}}. \quad (2.21)$$

The set of all functions  $f \in X^p$  for which  $\|f\|_{\gamma,p} < \infty$  is denoted by  $W_{\gamma,p}$ .

The following theorem [12, Chapter 7, Theorem 9.2], [57, Chapter III, Section 3.2, Corollary 7] shows the close connection between the apparently somewhat artificially

defined class  $W_{\gamma,p}$ , the rate of decrease of the degrees of approximation, and the smoothness in the classical sense.

**Theorem 2.3** *Let  $1 \leq p \leq \infty$ ,  $f \in X^p$ ,  $\gamma > 0$ ,  $r > \gamma$  be an integer,  $\gamma = s + \beta$  where  $s \geq 0$  is an integer chosen so that  $0 < \beta \leq 1$ , and let  $h$  be a twice continuously differentiable low pass filter.*

(a)  $f \in W_{\gamma,p}$  if and only if  $E_{n,p}(f) = \mathcal{O}(n^{-\gamma})$ . More precisely,

$$\|f\|_{\gamma,p} \sim \sup_{n \in \mathbb{N}} n^\gamma E_{n,p}(f) \sim \sup_{n \in \mathbb{N}} n^\gamma \|f - \sigma_n(h, f)\|_p, \quad (2.22)$$

where the constants involved in “ $\sim$ ” are independent of  $f$ .

(b) We have  $f \in W_{\gamma,p}$  if and only if  $f^{(s)} \in W_{\beta,p}$ .

We remark that since the quantity  $E_{n,p}(f)$  does not depend on the choice of  $r$  in the definition of  $W_{\gamma,p}$  except that  $r > \gamma$ , the class  $W_{\gamma,p}$  does not depend on the choice of  $r$  either, as long as  $r > \gamma$ . Second, we note that in the above theorem  $s \neq \lfloor \gamma \rfloor$ . In particular, for the function  $x \mapsto |\cos x|$ , one has  $\gamma = 1$ ,  $s = 0$  and  $\beta = 1$ . There are characterizations of the  $K$ -functional that are given directly in terms of the function  $f$  rather than as a regularization functional. For instance (cf. [57, Chapter III, Section 1.2, Theorem 6], [12, Chapter 6, Theorem 2.4]),

$$K_{r,p}(f, \delta) \sim \sup_{|t| \leq \delta} \left\| \sum_{k=0}^r (-1)^k \binom{r}{k} f(\circ + kt) \right\|_p,$$

where the right hand expression is called the  $r$ -th order modulus of smoothness of  $f$ , and the constants involved in “ $\sim$ ” are independent of  $f$  and  $\delta$ . In particular, if  $0 < \beta < 1$ , then  $f \in W_{\beta,\infty}$  is equivalent to the Lipschitz–Hölder condition on  $f$  ([57, Chapter III, Section 3.2, Corollary 7] [12, Chapter 7, Theorem 3.3]):

$$|f(x+t) - f(x)| \leq c \|f\|_{\beta,\infty} |t|^\beta, \quad x \in [-\pi, \pi]. \quad (2.23)$$

In modern approximation theory, it has become more customary to take the  $K$ -functional itself as a measurement of smoothness in various situations, taking such direct relationships for granted.

We end this subsection by pointing out an interesting property of the summability operator, referred to in approximation theory literature as realization of the  $K$ -functional.

**Theorem 2.4** *Let  $r \in \mathbb{N}$ ,  $1 \leq p \leq \infty$ , and  $f \in X^p$ . Let  $h$  be a twice continuously differentiable low pass filter. Then for  $n \in \mathbb{N}$ ,*

$$\|f - \sigma_n(h, f)\|_p + n^{-r} \|\sigma_n^{(r)}(h, f)\|_p \sim K_{r,p}(f, 1/n), \quad (2.24)$$

where the constants involved in “ $\sim$ ” are independent of both  $f$  and  $n$ .



Thus, when one is not interested in finding the function  $g$  that achieves the infimum in the definition of the  $K$ -functional,  $\sigma_n(h, f)$  supplies a near-optimal solution. Moreover, while finding the minimizer  $g$  is a separate optimization problem for each  $r$  and  $p$ , the summability operator works for *every*  $r$  and *every*  $p$ , and its construction does not involve any optimization.

The essential ideas behind the proof of Theorem 2.4 are in the paper [8] of Czipser and Freud. For a lack of easy reference, we include the simple proof.

*Proof of Theorem 2.4* The definition (2.20) of the  $K$ -functional shows that

$$K_{r,p}(f, 1/n) \leq \|f - \sigma_n(h, f)\|_p + n^{-r} \|\sigma_n^{(r)}(h, f)\|_p.$$

We prove the inequality in the other direction. Let  $g \in X^p$  be any function with  $g^{(r)} \in X^p$ . Then using Theorem 2.1 and the Favard estimate (2.10), we obtain

$$\begin{aligned} \|f - \sigma_n(h, f)\|_p &\leq \|f - g - \sigma_n(h, f - g)\|_p + \|g - \sigma_n(h, g)\|_p \leq \|f - g\|_p \\ &\quad + \|\sigma_n(h, f - g)\|_p + cE_{n/2}(g) \leq c \left\{ \|f - g\|_p + \frac{1}{n^r} \|g^{(r)}\|_p \right\}. \end{aligned} \quad (2.25)$$

Further, using the fact that  $\sigma_n(h, g)^{(r)} = \sigma_n(h, g^{(r)})$  and the Bernstein inequality (2.9), we deduce that

$$\begin{aligned} n^{-r} \|\sigma_n^{(r)}(h, f)\|_p &\leq n^{-r} \|\sigma_n(h, f - g)^{(r)}\|_p + n^{-r} \|\sigma_n(h, g)^{(r)}\|_p \\ &\leq c \{ \|\sigma_n(h, f - g)\|_p + n^{-r} \|\sigma_n(g^{(r)})\|_p \} \leq c \left\{ \|f - g\|_p + \frac{1}{n^r} \|g^{(r)}\|_p \right\}. \end{aligned}$$

Together with (2.25), we have shown that

$$\|f - \sigma_n(h, f)\|_p + n^{-r} \|\sigma_n^{(r)}(h, f)\|_p \leq c \left\{ \|f - g\|_p + \frac{1}{n^r} \|g^{(r)}\|_p \right\}.$$

Since  $g$  is an arbitrary function with  $g^{(r)} \in X^p$ , the definition (2.20) shows that

$$\|f - \sigma_n(h, f)\|_p + n^{-r} \|\sigma_n^{(r)}(h, f)\|_p \leq cK_{r,p}(f, 1/n).$$

□

## 2.3 Wavelet-like representation

In this section,  $h$  will denote a fixed, infinitely differentiable low pass filter. We have seen that  $f \in L^1$  if and only if  $\sigma_n(h, f) \rightarrow f$  in  $L^1$ . Since  $\sigma_n(h, f)$  is defined entirely in terms of the sequence  $\{\hat{f}(k)\}_{k \in \mathbb{Z}}$ , it follows that the sequence of Fourier coefficients of an integrable function determines the function uniquely. For this reason, this sequence is called the *frequency domain description* of  $f$ , while a formula giving

$f$  directly as a function of its argument is called a *space (or time) domain description*. There are some problems with the frequency domain description.

Except in the case of the  $L^2$  norm, where the Parseval identity is available, the frequency domain description of a function does not reveal its smoothness directly. For example, we consider

$$f_1(x) \stackrel{\text{def}}{=} \frac{\cos((\pi - x)/4)}{\sqrt{2} \sin(x/2)},$$

for which the formal Fourier series expansion is given by

$$1 + \frac{2}{\sqrt{\pi}} \sum_{k=1}^{\infty} \frac{\Gamma(k + 1/2)}{k!} \cos(kx),$$

see, e.g., [79, Chapter V, Formula (2.3)], so that, using Stirling's formula,  $\hat{f}_1(k) = \mathcal{O}(k^{-1/2})$ . The function  $f_1$  is discontinuous at 0. On the other hand, for

$$f_2(x) \stackrel{\text{def}}{=} \frac{1}{2} + \sum_{j=1}^{\infty} \frac{\cos(4^j x)}{2^j},$$

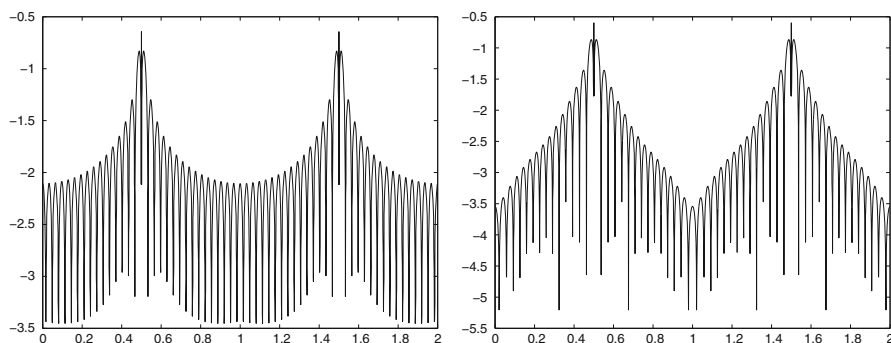
it is easy to verify using Theorem 2.3 that  $f_2 \in W_{1/2, \infty}$  even though  $\hat{f}_2(k) = \mathcal{O}(k^{-1/2})$  as well. For the function

$$f_3(x) \stackrel{\text{def}}{=} |\cos x|^{1/2} = \frac{\Gamma(3/2)}{\sqrt{2}\Gamma(5/4)^2} + \sum_{j=1}^{\infty} (-1)^j \frac{\sqrt{2}\Gamma(3/2)}{\Gamma(-1/4)\Gamma(5/4)} \frac{\Gamma(j-1/4)}{\Gamma(j+3/4)} \cos(2jx),$$

see, e.g., [17, p. 12, Eqn. (30)], the formulation (2.23) of the class  $W_{1/2, \infty}$  shows that  $f_3 \in W_{1/2, \infty}$ , but  $f_3 \notin W_{\gamma, \infty}$  for any  $\gamma > 1/2$ . However,  $\hat{f}_3(k) \sim k^{-3/2}$ . Finally,  $f_2$  is nowhere differentiable, while  $f_3$  and  $f_1$  both admit analytic continuations at all but finitely many points on  $\mathbb{T}$ .

In the last couple of decades, wavelet analysis has become popular as an alternative to Fourier series where the coefficients characterize local smoothness of the target function, see [10, Theorems 9.2.1, 9.2.2]. We find it interesting both from a theoretical as well as practical point of view to develop a similar expansion which can be computed using the classical Fourier coefficients, but achieves the same purpose. In this section, we review some of our results in this direction, sketching some proofs related to the summability kernel. This section is based on [58–60].

We pause again to discuss an example from [34] for the utility of the summability kernel and its localization. In Fig. 3 below, we report the log-plot of the error between  $x \mapsto |\cos x|^{1/4}$  and its Fourier projection of degree 31, compared to the error obtained by using our summability operator with a smooth filter  $h$ . In keeping with the converse theorem, the maximum error is of the same order of magnitude in both the cases, but it is clear that the error using the summability operator decreases rapidly as  $|x - \pi/2|$



**Fig. 3** The plot of the logarithm (base 10) of the absolute error between the function  $x \mapsto |\cos x|^{1/4}$ , and (left) its Fourier projection (right) trigonometric polynomial obtained by our summability operator, where the Fourier coefficients are estimated by 128 point FFT. The order of the trigonometric polynomials is 31 in each case. The numbers on the  $x$  axis are in multiples of  $\pi$ , the actual absolute errors are  $10^y$

or  $|x + \pi/2|$  increases. Thus, the summability operator is more robust with respect to local “singularities” in a function.

To resume our main discussion, we note that unlike the notion of derivatives, membership in  $W_{\gamma,p}$  is not a local property. The following is a standard way to define such spaces locally. We refer to an open connected subset of  $\mathbb{T}$  as an *arc*. Essentially, an arc is an open subinterval of  $[-\pi, \pi]$ , except that an arc might also be of the form  $[-\pi, a) \cup (b, \pi]$ . The class of all infinitely differentiable functions supported on an arc  $I$  will be denoted by  $C_I^\infty$ .

**Definition 2.3** Let  $1 \leq p \leq \infty$ ,  $\gamma > 0$ ,  $x_0 \in \mathbb{T}$ , and  $f \in L^1$ . We say that  $f \in W_{\gamma,p}(x_0)$  if there exists an arc  $I \ni x_0$  such that for every  $\phi \in C_I^\infty$ ,  $\phi f \in W_{\gamma,p}$ .

Thus, among the three functions listed above,  $f_1 \in W_{\gamma,\infty}(x_0)$  for all  $\gamma > 0$  and  $x_0 \neq 0$ ,  $f_2 \notin W_{\gamma,\infty}(x_0)$  for any  $x_0$  if  $\gamma > 1/2$ , and  $f_3 \in W_{\gamma,\infty}(x_0)$  for all  $\gamma > 0$ ,  $x_0 \neq \pm\pi/2$ .

We define the frame operators

$$\tau_j(h, f) \stackrel{\text{def}}{=} \begin{cases} \sigma_1(h, f), & \text{if } j = 0, \\ \sigma_{2j}(h, f) - \sigma_{2j-1}(h, f), & \text{if } j \in \mathbb{N}. \end{cases} \quad (2.26)$$

Note that these are the same as the polynomials  $P_j$  in (2.11). We let

$$\Psi_j(h, t) \stackrel{\text{def}}{=} \Phi_{2j+1}(h, t) - \Phi_{2j-2}(h, t), \quad j = 0, 1, \dots, \quad (2.27)$$

where we recall our convention that  $\Phi_n(h, t) = 0$  if  $n < 1$ .

We will not use the following constructions in the rest of the paper, but state them for the sake of completeness. Let

$$g^*(t) \stackrel{\text{def}}{=} \sqrt{h(t) - h(2t)}. \quad (2.28)$$

We write for  $t \in \mathbb{T}$ , and  $f \in L^1$ ,

$$\Psi_j^*(t) \stackrel{\text{def}}{=} \begin{cases} \Phi_{2^j}(g^*, t), & j \in \mathbb{N}, \\ 1, & j = 0, \end{cases} \quad (2.29)$$

and

$$\tau_j^*(h, f) \stackrel{\text{def}}{=} \sigma_{2^j}(g^*, f), \quad j = 1, 2, \dots, \quad \tau_0^*(h, f, t) \stackrel{\text{def}}{=} \hat{f}(0). \quad (2.30)$$

Finally, we set

$$v_{k,j} = k\pi/2^j, \quad k = -2^j + 1, \dots, 2^j, \quad j = 0, 1, 2, \dots \quad (2.31)$$

It can be verified using the formula for the sum of geometric series that

$$\frac{1}{2^{j+1}} \sum_{k=-2^j+1}^{2^j} \exp(i\ell v_{k,j}) = \begin{cases} 0, & \text{if } \ell = \pm 1, \dots, \pm(2^{j+1} - 1), \\ 1, & \text{if } \ell = 0, \end{cases}$$

and hence, the following quadrature formula holds.

$$\frac{1}{2\pi} \int_{\mathbb{T}} P(t) dt = \frac{1}{2^{j+1}} \sum_{k=-2^j+1}^{2^j} P(v_{k,j}), \quad P \in \mathbb{H}_{2^{j+1}}. \quad (2.32)$$

The following theorem states the wavelet-like (frame) expansion of functions in  $X^p$ . We note that, unlike classical Littlewood–Paley expansions or wavelet expansions, our results are valid also for  $f \in X^p$ , where  $p = 1$  and  $p = \infty$  are included. We note that the coefficients  $\tau_j(h, f, v_{k,j})$  and  $\tau_j^*(h, f, v_{k,j})$  can be computed as finite linear combinations of the Fourier coefficients of  $f$ .

**Theorem 2.5** *Let  $1 \leq p \leq \infty$ ,  $\gamma > 0$ ,  $f \in X^p$ , and let  $h$  be an infinitely differentiable low pass filter.*

(a) *We have*

$$\begin{aligned} f &= \sum_{j=0}^{\infty} \tau_j(h, f) = \sum_{j=0}^{\infty} \frac{1}{2^{j+1}} \sum_{k=-2^j+1}^{2^j} \tau_j(h, f, v_{k,j}) \Psi_j(h, \circ - v_{k,j}) \\ &= \sum_{j=0}^{\infty} \frac{1}{2^{j+1}} \sum_{k=-2^j+1}^{2^j} \tau_j^*(h, f, v_{k,j}) \Psi_j^*(h, \circ - v_{k,j}), \end{aligned} \quad (2.33)$$

*with convergence in the sense of  $X^p$ .*

(b) If  $f \in L^2$  then

$$\|f\|_2^2 \leq 2 \sum_{j=0}^{\infty} \|\tau_j(h, f)\|_2^2 = \sum_{j=0}^{\infty} \frac{1}{2^j} \sum_{k=-2^j+1}^{2^j} |\tau_j(h, f, v_{k,j})|^2 \leq 2 \|f\|_2^2 \quad (2.34)$$

and

$$\|f\|_2^2 = \sum_{j=0}^{\infty} \|\tau_j^*(h, f)\|_2^2 = \sum_{j=0}^{\infty} \frac{1}{2^{j+1}} \sum_{k=-2^j+1}^{2^j} |\tau_j^*(h, f, v_{k,j})|^2. \quad (2.35)$$

*Proof* The first equation in (2.33) follows from (2.8) by a telescoping series argument, cf. the proof of (2.13) (the polynomials denoted in the latter by  $P_j$  are in fact  $\tau_j(h, f)$ ). A comparison of Fourier coefficients shows that for  $j = 0, 1, \dots$ ,

$$\begin{aligned} \tau_j(h, f, x) &= \frac{1}{2\pi} \int_{\mathbb{T}} \tau_j(h, f, t) \Psi_j(h, x - t) dt \\ &= \frac{1}{2\pi} \int_{\mathbb{T}} \tau_j^*(h, f, t) \Psi_j^*(h, x - t) dt. \end{aligned}$$

The quadrature formula (2.32) can be used to express  $\tau_j(h, f)$  in the first equation in (2.33) as the inner sums as indicated in the remaining two equations there.

We observe that  $h(t) - h(2t) \geq 0$  for all  $t$  and  $h(t) - h(2t) \neq 0$  only if  $1/4 < t < 1$ . So, for any  $k \in \mathbb{Z} \setminus \{0\}$ , if  $m_k$  is the integer part of  $\log_2 |k|$ , then

$$h\left(\frac{k}{2^j}\right) - h\left(\frac{k}{2^{j-1}}\right) \neq 0 \text{ only if } j = m_k + 1 \text{ or } j = m_k + 2$$

Consequently, recalling that  $0 \leq h(t) \leq 1$  for all  $t \in \mathbb{R}$ , we obtain for  $k \in \mathbb{Z} \setminus \{0\}$  that

$$\begin{aligned} 1 &= \left( \sum_{j=1}^{\infty} h\left(\frac{k}{2^j}\right) - h\left(\frac{k}{2^{j-1}}\right) \right)^2 \leq 2 \sum_{j=1}^{\infty} \left( h\left(\frac{k}{2^j}\right) - h\left(\frac{k}{2^{j-1}}\right) \right)^2 \\ &\leq 2 \sum_{j=1}^{\infty} h\left(\frac{k}{2^j}\right) - h\left(\frac{k}{2^{j-1}}\right) = 2. \end{aligned} \quad (2.36)$$

Using the Parseval identity and Fubini's theorem, we see that

$$\begin{aligned} \sum_{j=0}^{\infty} \|\tau_j(h, f)\|_2^2 &= |\hat{f}(0)|^2 + \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z} \setminus \{0\}} \left( h\left(\frac{k}{2^j}\right) - h\left(\frac{k}{2^{j-1}}\right) \right)^2 |\hat{f}(k)|^2 \\ &= |\hat{f}(0)|^2 + \sum_{k \in \mathbb{Z} \setminus \{0\}} |\hat{f}(k)|^2 \sum_{j=1}^{\infty} \left( h\left(\frac{k}{2^j}\right) - h\left(\frac{k}{2^{j-1}}\right) \right)^2. \end{aligned}$$

Using the Parseval identity again and (2.36), we conclude that

$$\|f\|_2^2 = \sum_{k \in \mathbb{Z}} |\hat{f}(k)|^2 \leq 2 \sum_{j=0}^{\infty} \|\tau_j(h, f)\|_2^2 \leq 2\|f\|_2^2. \quad (2.37)$$

The quadrature formula (2.32) shows that

$$\|\tau_j(h, f)\|_2^2 = \frac{1}{2^{j+1}} \sum_{k=-2^j+1}^{2^j} |\tau_j(h, f, v_{k,j})|^2, \quad j = 0, 1, \dots$$

Together with (2.37), this completes the proof of (2.34).

Another simple application of the Parseval identity gives

$$\|f\|_2^2 = \sum_{j=0}^{\infty} \|\tau_j^*(h, f)\|_2^2,$$

which, together with the quadrature formula (2.32) leads to (2.35).  $\square$

The following theorem shows that the coefficients of the expansions in (2.33) provide a complete characterization of local smoothness classes, analogous to the corresponding theorems in wavelet analysis.

**Theorem 2.6** *Let  $1 \leq p \leq \infty$ ,  $\gamma > 0$ ,  $x_0 \in \mathbb{T}$ ,  $f \in X^p$ , and let  $h$  be an infinitely differentiable low pass filter. Then the following statements are equivalent.*

- (a)  $f \in W_{\gamma,p}(x_0)$ .
- (b) *There exists an arc  $I \ni x_0$  such that*

$$\left\{ \frac{1}{2^{j+1}} \sum_{k:v_{k,j} \in I} |\tau_j(h, f, v_{k,j})|^p \right\}^{1/p} = \mathcal{O}(2^{-j\gamma}), \quad 1 \leq p < \infty. \quad (2.38)$$

*In the case when  $p = \infty$ , the above estimate is interpreted as*

$$\max_{k:v_{k,j} \in I} |\tau_j(h, f, v_{k,j})| = \mathcal{O}(2^{-j\gamma}). \quad (2.39)$$

- (c) *There exists an arc  $I$  containing  $x_0$  such that*

$$\left\{ \frac{1}{2^{j+1}} \sum_{k:v_{k,j} \in I} |\tau_j^*(h, f, v_{k,j})|^p \right\}^{1/p} = \mathcal{O}(2^{-j\gamma}), \quad (2.40)$$

*with a modification as done in (b) in the case when  $p = \infty$ .*

We will prove only the equivalence of parts (a) and (b) above; the equivalence between parts (b) and (c) do not add much to the concepts we wish to emphasize here. This proof depends on the following Marcinkiewicz–Zygmund–type inequalities, see [79, Chapter X, Theorem (7.5) and the remark thereafter, and Theorem (7.28)].

**Lemma 2.1** For  $1 \leq p \leq \infty$  and  $T \in \mathbb{H}_{2j+1}$ , we have

$$\|T\|_p \sim \left\{ \frac{1}{2^{j+1}} \sum_{k=-2^j+1}^{2^j} |T(v_{k,j})|^p \right\}^{1/p}, \quad (2.41)$$

where the usual interpretation for the middle expression is assumed when  $p = \infty$ , and the constants involved in “ $\sim$ ” are independent of  $j$  and  $T$ .

*Proof of (a)  $\iff$  (b) of Theorem 2.6* In this proof, we will write for any arc  $I$ , integer  $j \geq 1$  and  $g : \mathbb{T} \rightarrow \mathbb{R}$ ,

$$[g]_{j,p,I} = \left\{ \frac{1}{2^{j+1}} \sum_{k: v_{k,j} \in I} |g(v_{k,j})|^p \right\}^{1/p}, \quad 1 \leq p \leq \infty.$$

Let  $f \in W_{\gamma,p}(x_0)$ , and  $J \ni x_0$  be an arc such that  $\phi f \in W_{p,\gamma}$  for every infinitely differentiable function  $\phi$  supported on  $J$ . We consider subarcs  $I \subset I' \subset J$  with  $x_0 \in I$ , and an infinitely differentiable function  $\psi$  which is equal to 1 on  $I'$  and 0 outside  $J$ . We choose and fix an integer  $S > \max(1, \gamma)$ . Since  $h$  is infinitely differentiable, it follows from (2.6) that

$$|\Psi_j(t)| \leq c2^j \min \left( 1, (2^j|t|)^{-S-1} \right), \quad t \in \mathbb{T}, \quad j \geq 0. \quad (2.42)$$

Therefore, for  $x \in I$ ,

$$\begin{aligned} |\tau_j((1 - \psi)f, x)| &\leq \frac{1}{2\pi} \int_{\mathbb{T}} |(1 - \psi(t))f(t)\Psi_j(x - t)| dt \\ &= \frac{1}{2\pi} \int_{\mathbb{T} \setminus I'} |(1 - \psi(t))f(t)\Psi_j(x - t)| dt \leq c(I, I')2^{-jS} \|f\|_1, \end{aligned}$$

and

$$|\tau_j(f, x)| \leq |\tau_j(\psi f, x)| + |\tau_j((1 - \psi)f, x)| \leq |\tau_j(\psi f, x)| + c(I, I')2^{-jS} \|f\|_1.$$

Consequently, using Lemma 2.1, we conclude that

$$\begin{aligned} [\tau_j(f)]_{j,p,I} &\leq [\tau_j(\psi f)]_{j,p,I} + c(I, I')2^{-jS} \|f\|_1 \\ &\leq [\tau_j(\psi f)]_{j,p,\mathbb{T}} + c(I, I')2^{-jS} \|f\|_1 \\ &\leq c(I, I')\{\|\tau_j(\psi f)\|_p + 2^{-jS} \|f\|_1\}. \end{aligned} \quad (2.43)$$

Since  $\tau_j(P) \equiv 0$  for all  $P \in \mathbb{H}_{2j-1}$ , the boundedness of the operators  $\tau_j$  implies that for any  $P \in \mathbb{H}_{2j-1}$ ,

$$\|\tau_j(\psi f)\|_p = \|\tau_j(\psi f - P)\|_p \leq c\|\psi f - P\|_p,$$

and, hence,

$$\|\tau_j(\psi f)\|_p \leq cE_{2j-1,p}(\psi f).$$

Therefore, (2.43) shows that

$$[\tau_j(f)]_{j,p,I} \leq c(I, I')\{E_{2j-1,p}(\psi f) + 2^{-jS}\|f\|_1\}.$$

Since  $\psi f \in W_{\gamma,p}$  and  $S > \gamma$ , an application of Theorem 2.3 now leads to part (b).

Conversely, let  $I \ni x_0$  be an arc such that

$$[\tau_j(f)]_{j,p,I} = \mathcal{O}(2^{-j\gamma}), \quad (2.44)$$

and let  $\phi$  be any infinitely differentiable function supported on  $I$ . The Favard estimate then shows that for every  $j \geq 1$  there exists  $R_j \in \mathbb{H}_{2j}$  such that

$$\|\phi - R_j\|_\infty \leq c(\phi)2^{-jS}. \quad (2.45)$$

Therefore, in view of (2.7), we obtain

$$\begin{aligned} E_{2j+1,p}(\phi f) &\leq \|\phi f - R_j \sigma_{2j}(h, f)\|_p \leq \|\phi(f - \sigma_{2j}(h, f))\|_p + \|(\phi - R_j)\sigma_{2j}(h, f)\|_p \\ &\leq \|\phi(f - \sigma_{2j}(h, f))\|_p + c(\phi)2^{-jS}\|f\|_p. \end{aligned} \quad (2.46)$$

Using (2.33), Lemma 2.1, (2.45), and (2.7) in that order, we see that

$$\begin{aligned} \|\phi(f - \sigma_{2j}(h, f))\|_p &= \left\| \phi \sum_{k=j+1}^{\infty} \tau_k(h, f) \right\|_p \leq \sum_{k=j+1}^{\infty} \|\phi \tau_k(h, f)\|_p \\ &\leq \sum_{k=j+1}^{\infty} \|R_k \tau_k(h, f)\|_p + \sum_{k=j+1}^{\infty} \|(\phi - R_k) \tau_k(h, f)\|_p \\ &\leq c \left\{ \sum_{k=j+1}^{\infty} [R_k \tau_k(h, f)]_{k,p,\mathbb{T}} + \sum_{k=j+1}^{\infty} \|(\phi - R_k) \tau_k(h, f)\|_p \right\} \\ &\leq c \left\{ \sum_{k=j+1}^{\infty} [\phi \tau_k(h, f)]_{k,p,\mathbb{T}} + \sum_{k=j+1}^{\infty} [(\phi - R_k) \tau_k(h, f)]_{k,p,\mathbb{T}} + \sum_{k=j+1}^{\infty} \|(\phi - R_k) \tau_k(h, f)\|_p \right\} \\ &\leq c(\phi) \left\{ \sum_{k=j+1}^{\infty} [\tau_k(h, f)]_{k,p,I} + \sum_{k=j+1}^{\infty} 2^{-kS} [\tau_k(h, f)]_{k,p,\mathbb{T}} + \sum_{k=j+1}^{\infty} 2^{-kS} \|\tau_k(h, f)\|_p \right\} \end{aligned}$$



$$\begin{aligned} &\leq c(\phi) \left\{ \sum_{k=j+1}^{\infty} [\tau_k(h, f)]_{k,p,I} + \sum_{k=j+1}^{\infty} 2^{-kS} \|\tau_k(h, f)\|_p \right\} \\ &\leq c(\phi) \left\{ \sum_{k=j+1}^{\infty} [\tau_k(h, f)]_{k,p,I} + \sum_{k=j+1}^{\infty} 2^{-kS} \|f\|_p \right\}. \end{aligned}$$

Thus, the assumption (2.44) leads to

$$\|\phi(f - \sigma_{2^j}(h, f))\|_p = \mathcal{O}(2^{-j\gamma}).$$

Since  $S > \gamma$ , this estimate and (2.46) show that  $E_{2^{j+1},p}(\phi f) = \mathcal{O}(2^{-j\gamma})$ . Therefore, Theorem 2.3 implies that for every infinitely differentiable  $\phi$  supported on  $I$  we have  $\phi f \in W_{\gamma,p}$ , that is, (a) holds.  $\square$

We note that an expansion of the form (2.33) is used often in approximation theory. We reserve the term *wavelet-like representation* to indicate that the behavior of the terms of the expansion characterize **local smoothness** of the target function. Thus, for example, although the expansion

$$f = \sum_{j=0}^{\infty} (v_{2^{j+1}}(f) - v_{2^j}(f)) + v_1(f)$$

is very similar to (2.33), and holds for every  $f \in X^p$   $p \in [1, \infty]$ , we do not refer to this expansion as a wavelet-like representation, because the localization properties of the operators  $v_n$  are not strong enough to admit an analogue of Theorem 2.6 for characterization of local smoothness classes for the smoothness parameter  $> 1$ .

### 3 Multivariate analogues

#### 3.1 Notation

In what follows,  $q \geq 2$  is a fixed integer and the various constants will depend on  $q$ . As usual, the notation  $\mathbb{T}^q$  denotes the  $q$  dimensional torus, that is, a  $q$ -fold cross product of  $\mathbb{T}$  with itself. As in the univariate case, functions on  $\mathbb{T}^q$  can be considered as functions on  $\mathbb{R}^q$ ,  $2\pi$ -periodic in each variable. An *arc* in this context has the form  $\prod_{j=1}^q I_k$ , where each  $I_k \subseteq \mathbb{T}$  is a univariate arc as defined in Sect. 2. An element of  $\mathbb{R}^q$  will be denoted in bold face, e.g.,  $\mathbf{x} = (x_1, \dots, x_q)$ . We find it convenient to write  $x$  also for the vector  $(x, x, \dots, x)$ ; e.g.,  $0$  denotes both the scalar  $0$  and the vector  $(0, 0, \dots, 0)$ . We hope that it will be clear from the context whether a scalar is intended or a vector with equal components is intended. When applied to vectors, univariate operations and relations will be interpreted in a coordinatewise sense; e.g.,  $\mathbf{x} \geq 0$  means that  $x_j \geq 0$  for all  $j = 1, 2, \dots, q$ ,  $\mathbf{x}^{\mathbf{y}} = (x_1^{y_1}, \dots, x_q^{y_q})$ , whenever the expressions are defined, and so forth. The notation  $\mathbf{x} \cdot \mathbf{y}$  denotes the inner product

between  $\mathbf{x}$  and  $\mathbf{y}$ . For  $0 < p \leq \infty$ , we define

$$|\mathbf{x}|_p = \begin{cases} (\sum_{k=1}^q |x_k \bmod (2\pi)|^p)^{1/p}, & \text{if } 0 < p < \infty, \\ \max_{1 \leq k \leq q} |x_k \bmod (2\pi)|, & \text{if } p = \infty. \end{cases}$$

In general, many of the definitions of norms and related expressions, are very similar to those in the univariate case, and we will often omit the dimension  $q$  whenever we feel that writing it explicitly makes the notation unnecessarily cumbersome.

If  $f : \mathbb{T}^q \rightarrow \mathbb{R}$  is differentiable, we write  $\partial_r f$  to denote the partial derivative of  $f$  with respect to the  $r$ -th variable. If  $\mathbf{r} \in \mathbb{Z}_+^q$  and  $f$  is sufficiently smooth, we write  $\partial^{\mathbf{r}} f$  to denote the partial derivative indicated by  $\mathbf{r}$ .

The measure  $\mu_q^*$  denotes the Lebesgue measure on  $\mathbb{T}^q$  normalized to 1. If  $f : \mathbb{T}^q \rightarrow \mathbb{C}$  is Lebesgue measurable, and  $A \subseteq \mathbb{T}^q$  is Lebesgue measurable, we write

$$\|f\|_{p,A} \stackrel{\text{def}}{=} \begin{cases} \left\{ \int_A |f(\mathbf{x})|^p d\mu_q^*(\mathbf{x}) \right\}^{1/p}, & \text{if } 0 < p < \infty, \\ \mu_q^*\text{-ess sup}_{\mathbf{x} \in A} |f(\mathbf{x})|, & \text{if } p = \infty. \end{cases} \quad (3.1)$$

As before, if  $A = \mathbb{T}^q$ , then it will be omitted from the notation; e.g.,  $\|f\|_p = \|f\|_{p,\mathbb{T}^q}$ . The spaces  $L^p$  are defined as usual.

### 3.2 Approximation theory

There are many ways to define trigonometric polynomials of several variables. In particular, for  $p$  such that  $0 < p \leq \infty$ , a trigonometric polynomial of  $\ell^p$ -degree  $< n$  is a function of the form

$$\mathbf{x} \mapsto \sum_{\mathbf{k}: |\mathbf{k}|_p < n} a_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{x}).$$

When  $p = \infty$ , we refer to *coordinatewise-degree*  $< n$ , whereas if  $p = 1$ , we refer to *total degree*  $< n$ , and when  $p = 2$ , we refer to the *spherical degree*  $< n$ . The dimensions of all these spaces are always  $\mathcal{O}(n^q)$ . In our discussion below, we will restrict ourselves to the class  $\mathbb{H}_n^q$  of trigonometric polynomials of spherical degree  $< n$ , but the results will be equally valid also for other values of  $p$ . The  $L^p$  closure of  $\cup_{n \geq 0} \mathbb{H}_n^q$  will be denoted by  $X^p$  (or  $X^p(\mathbb{T}^q)$  if some confusion is likely to result).

If  $f \in X^p(\mathbb{T}^q)$  and  $n \geq 0$ , then the degree of approximation of  $f$  from  $\mathbb{H}_n^q$  is defined by

$$E_{n,p}(f) \stackrel{\text{def}}{=} E_{q;n,p}(f) \stackrel{\text{def}}{=} \inf \{ \|f - P\|_p : P \in \mathbb{H}_n^q \}.$$

Again, the symbol  $q$  will be omitted when we don't expect any confusion. If  $f \in L^1$ , its Fourier coefficient is defined by

$$\hat{f}(\mathbf{k}) = \int_{\mathbb{T}^q} f(\mathbf{t}) \exp(i\mathbf{k} \cdot \mathbf{t}) d\mu_q^*(\mathbf{t}), \quad \mathbf{k} \in \mathbb{Z}^q.$$

Let  $H : \mathbb{R}^q \rightarrow \mathbb{R}$  be compactly supported. The role of the kernel  $\Phi_n$  is played by

$$\Phi_n(H, \mathbf{t}) \stackrel{\text{def}}{=} \Phi_{n,q}(H, \mathbf{t}) \stackrel{\text{def}}{=} \sum_{\mathbf{k} \in \mathbb{Z}^q} H\left(\frac{\mathbf{k}}{n}\right) \exp(i\mathbf{k} \cdot \mathbf{t}), \quad n \in \mathbb{N}, \mathbf{t} \in \mathbb{T}^q,$$

and the corresponding operator is defined by

$$\begin{aligned} \sigma_n(H, f, \mathbf{x}) &= \int_{\mathbb{T}^q} f(\mathbf{t}) \Phi_n(H, \mathbf{x} - \mathbf{t}) d\mu_q^*(\mathbf{t}) \\ &= \sum_{\mathbf{k} \in \mathbb{Z}^q} H\left(\frac{\mathbf{k}}{n}\right) \hat{f}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x}), \quad n \in \mathbb{N}, \mathbf{x} \in \mathbb{T}^q. \end{aligned}$$

Of particular interest is the case when  $H(\mathbf{t}) = h(|\mathbf{t}|_2)$  for some compactly supported  $h : [0, \infty) \rightarrow \mathbb{R}$ . We will overload the notation again, and denote the corresponding kernel and operator by  $\Phi_n(h, \circ)$  and  $f \mapsto \sigma_n(h, f)$ , with the domain of  $\Phi_n$  and  $f$  making it clear which meaning is intended. We observe that when the mapping  $\mathbf{x} \mapsto h(|\mathbf{x}|_2)$  is integrable on  $\mathbb{R}^q$ , then its Fourier transform is also *radial*, that is, it has the form  $\mathbf{x} \mapsto F(|\mathbf{x}|_2)$  for some  $F : [0, \infty) \rightarrow \mathbb{C}$  [75, Chapter IV, Theorem 3.3]. Therefore, if the Poisson summation formula holds, then  $\Phi_n(h, \circ)$  is a radial function,  $2\pi$ -periodic in each of its variables.

The analogue of Theorem 2.1 is the following, where the conditions on the filter  $H$  are a bit stronger to ensure the validity of the Poisson summation formula enlisted in the proof of the theorem.

**Theorem 3.1** *Let  $S > q$  be an integer and let  $H : \mathbb{R}^q \rightarrow \mathbb{R}$  be an  $S$ -times continuously differentiable, compactly supported function.*

(a) *We have*

$$|\Phi_n(H, \mathbf{t})| \leq c(H)n^q \min\left(1, (n|\mathbf{t}|_2)^{-S}\right), \quad \mathbf{t} \in \mathbb{T}, n \geq 1. \quad (3.2)$$

(b) *For  $1 \leq p \leq \infty$ ,*

$$\|\sigma_n(H, f)\|_p \leq c(H)\|f\|_p, \quad f \in L^p, n \in \mathbb{N}. \quad (3.3)$$

(c) *Let  $h : \mathbb{R} \rightarrow [0, 1]$  be an  $S$  times continuously differentiable low pass filter and  $n \in \mathbb{N}$ . Then  $\sigma_n(h, P) = P$  for all  $P \in \mathbb{H}_{n/2}^q$ . Moreover, for  $1 \leq p \leq \infty$  and  $f \in L^p$ ,*

$$E_{n,p}(f) \leq \|f - \sigma_n(h, f)\|_p \leq c(h)E_{n/2,p}(f). \quad (3.4)$$

*Proof* The proof of parts (a) and (b) are given in [5, Section 6.1]. To prove part (c), we let  $H_1(\mathbf{x}) = h(|\mathbf{x}|_2)$  and observe that since  $h$  is constant in a neighborhood of 0,

$$\nabla H_1(\mathbf{x}) = \frac{\mathbf{x}}{|\mathbf{x}|_2} h'(|\mathbf{x}|_2) = 0 \quad \text{in a neighborhood of 0.}$$

Since  $h$  is  $S$ -times continuously differentiable, it follows that  $H_1$  is as well. Since  $\sigma_n(h, f) = \sigma_n(H_1, f)$ , the estimate (3.3) holds with  $H_1$  in place of  $H$ , that is,  $\sigma_n(h, f)$  in place of  $\sigma_n(H, f)$ . The remainder of the proof is verbatim the same as the corresponding part of Theorem 2.1.  $\square$

We observe an important corollary of Theorem 3.1.

**Corollary 3.1** *If  $n \in \mathbb{N}$  and  $T \in \mathbb{H}_n^q$ , then for  $1 \leq p \leq \infty$  and  $1 \leq r \leq q$ ,*

$$\|\partial_r T\|_p \leq cn \|T\|_p. \quad (3.5)$$

*Proof* Since  $\widehat{\partial_r T}(\mathbf{k}) = ik_r \hat{T}(\mathbf{k})$  for  $\mathbf{k} \in \mathbb{Z}^q$ , we may use Theorem 3.1 with an appropriate smooth low pass filter  $h$ , and the function  $H(\mathbf{x}) = x_r h(|\mathbf{x}|_2)$  as in the proof of the univariate Bernstein inequality, see Theorem 2.2(a), to prove (3.5).

We will denote the Laplacian operator by  $\Delta \stackrel{\text{def}}{=} \sum_{j=1}^q \partial_j^2$ , and observe that for a sufficiently smooth  $f$ ,

$$\widehat{(-\Delta f)}(\mathbf{k}) = |\mathbf{k}|_2^2 \hat{f}(\mathbf{k}),$$

for each  $\mathbf{k} \in \mathbb{Z}^q$ . For  $r > 0$ , we define the differential operator  $(-\Delta)^{r/2}$  formally by

$$\widehat{(-\Delta)^{r/2} f}(\mathbf{k}) \stackrel{\text{def}}{=} |\mathbf{k}|_2^r \hat{f}(\mathbf{k}) \quad (3.6)$$

for  $f \in L^1$  and  $\mathbf{k} \in \mathbb{Z}^q$ . Clearly, if  $T \in \mathbb{H}_n^q$  for some  $n \in \mathbb{N}$ , then  $(-\Delta)^{r/2} T$  is well defined for all  $r > 0$ .

With respect to these operators, the analogous Favard estimate and Bernstein inequality are the following.

**Theorem 3.2** (a) *Let  $1 \leq p \leq \infty$  and let  $r$  be a positive even integer. Then for all  $f \in X^p$  for which  $(-\Delta)^{r/2} f \in X^p$ , we have*

$$E_{n,p}(f) \leq cn^{-r} \|(-\Delta)^{r/2} f\|_p.$$

(b) *Let  $T \in \mathbb{H}_n^q$  and let  $r$  be a positive even integer. Then for  $1 \leq p \leq \infty$ ,*

$$\|(-\Delta)^{r/2} T\|_p \leq cn^r \|T\|_p.$$

The  $K$ -functional appropriate for the multivariate theory is defined (with another overload of notation) as follows.

**Definition 3.1** If  $r \geq 1$  is an even integer,  $1 \leq p \leq \infty$ , and  $f \in X^p$ , then we define

$$K_{r,p}(f, \delta) \stackrel{\text{def}}{=} K_{q;r,p}(f, \delta) \stackrel{\text{def}}{=} \inf \{ \|f - g\|_p + \delta^r \|(-\Delta)^{r/2} g\|_p \}, \quad (3.7)$$

where the infimum is taken over all  $g$  for which  $(-\Delta)^{r/2} g \in X^p$ . If  $\gamma > 0$  and  $r > \gamma$  is an even integer, then we define

$$\|f\|_{\gamma,p} \stackrel{\text{def}}{=} \|f\|_{q;\gamma,p} \stackrel{\text{def}}{=} \sup_{0 < \delta < 1/2} \frac{K_{r,p}(f, \delta)}{\delta^\gamma}. \quad (3.8)$$

The set of all functions  $f \in X^p$  for which  $\|f\|_{\gamma,p} < \infty$  is denoted by  $W_{\gamma,p}$  or, if some confusion is likely to happen, then by  $W_{q;\gamma,p}$ .

The following theorem is the direct analogue of Theorem 2.3 and it is proved in the same way.

**Theorem 3.3** Let  $1 \leq p \leq \infty$ ,  $f \in X^p$ ,  $\gamma > 0$ ,  $r > \gamma$  be an even integer,  $\gamma = s + \beta$  where  $s \geq 0$  is an integer chosen so that  $0 < \beta \leq 1$ . Let  $S > q$  be an integer and  $h$  be an  $S$ -times continuously differentiable low pass filter.

(a)  $f \in W_{\gamma,p}$  if and only if  $E_{n,p}(f) = \mathcal{O}(n^{-\gamma})$ . More precisely,

$$\|f\|_{\gamma,p} \sim \sup_{n \in \mathbb{N}} n^\gamma E_{n,p}(f) \sim \sup_{n \in \mathbb{N}} n^\gamma \|f - \sigma_n(h, f)\|_p, \quad (3.9)$$

where the constants involved in “ $\sim$ ” are independent of  $f$ .

(b) We have  $f \in W_{\gamma,p}$  if and only if  $(-\Delta)^{s/2} f \in W_{\beta,p}$ .

Finally, we state the following theorem.

**Theorem 3.4** Let  $1 \leq p \leq \infty$ ,  $f \in X^p$ , and let  $r \geq 1$  be an even integer. Let  $S > q$  be an integer and let  $h$  be an  $S$ -times continuously differentiable low pass filter. Then for  $n \in \mathbb{N}$ ,

$$\|f - \sigma_n(h, f)\|_p + n^{-r} \|(-\Delta)^{r/2} \sigma_n(h, f)\|_p \sim K_{r,p}(f, 1/n), \quad (3.10)$$

where the constants involved in “ $\sim$ ” are independent of  $f$  and  $n$ .

### 3.3 Discretization

For many applications in learning theory, the information available about the target function  $f$  consists of its values  $f(\mathbf{y}_k)$  at finitely many points  $\{\mathbf{y}_k\}_{k=1}^M$  but one cannot prescribe in advance the precise location of these points. The goal of this section is to survey the ideas behind a construction of summability operators based on such information which have properties similar to those of the operators  $\sigma_n(h)$ . An essential

ingredient is to obtain real numbers  $w_k$  such that for an integer  $N \geq 0$  as high as possible, both the *quadrature formula* (3.12) and *M-Z (Marcinkiewicz-Zygmund) inequalities* (3.13) below hold. It will be shown in Theorem 3.5 that such weights can always be found with the desired degree  $N$  being dependent on the so-called *density content* of the set  $\{\mathbf{y}_k\}$  [43].

**Definition 3.2** Let  $\mathcal{C} \stackrel{\text{def}}{=} \{\mathbf{y}_k\}_{k=1}^M \subset \mathbb{T}^q$ . We define the *density content*  $\delta(\mathcal{C})$  and the *minimal separation*  $\eta(\mathcal{C})$  by

$$\delta(\mathcal{C}) = \max_{\mathbf{x} \in \mathbb{T}^q} \min_{1 \leq k \leq M} |\mathbf{x} - \mathbf{y}_k|_\infty, \quad \eta(\mathcal{C}) \stackrel{\text{def}}{=} \min_{1 \leq k \neq j \leq M} |\mathbf{y}_k - \mathbf{y}_j|_\infty. \quad (3.11)$$

We note that the density content has been referred to in the literature also as *fill distance* or *mesh norm*.

**Theorem 3.5** Let  $\mathcal{C} \stackrel{\text{def}}{=} \{\mathbf{y}_k\}_{k=1}^M \subset \mathbb{T}^q$  and  $\delta(\mathcal{C}) \leq 1$ . There exists a positive constant  $\alpha$  dependent only on  $q$  with the property that there are non-negative numbers  $\{w_k\}_{k=1}^M$  such that for  $N \leq \alpha\delta(\mathcal{C})^{-1}$  we have

$$\sum_{k=1}^M w_k P(\mathbf{y}_k) = \int_{\mathbb{T}^q} P(\mathbf{t}) d\mu_q^*(\mathbf{t}), \quad P \in \mathbb{H}_N^q, \quad (3.12)$$

and

$$\sum_{k=1}^M |w_k P(\mathbf{y}_k)| \sim \int_{\mathbb{T}^q} |P(\mathbf{t})| d\mu_q^*(\mathbf{t}), \quad P \in \mathbb{H}_N^q, \quad (3.13)$$

where the constants involved in “ $\sim$ ” depend only on  $q$  and not on  $\mathcal{C}$ ,  $M$ ,  $N$ , the choice of the weights  $w_k$ , or the polynomials  $P$ .

*Remark* Let  $\mathcal{C} \subset \mathbb{T}^q$  be a finite set. It is easy to verify that  $\eta(\mathcal{C}) \leq 4\delta(\mathcal{C})$ . When  $\delta(\mathcal{C}) \leq 1$ , we can always select a subset of  $\mathcal{C}$  for which the minimal separation and the density content have the same order of magnitude. Let  $m$  be the integer part of  $2\pi/(3\delta(\mathcal{C}))$ . For a multi-integer  $\mathbf{k}$  with  $0 \leq \mathbf{k} \leq m - 1$ , we define

$$I_{\mathbf{k}} = I_{\mathbf{k}, \mathcal{C}} \stackrel{\text{def}}{=} \prod_{j=1}^q \left[ -\pi + \frac{2k_j\pi}{m}, -\pi + \frac{2(k_j + 1)\pi}{m} \right]. \quad (3.14)$$

Then  $\{I_{\mathbf{k}}\}$ 's are mutually disjoint arcs except for common boundaries of measure 0 and their union is  $\mathbb{T}^q$ . Let  $\mathbf{z}_{\mathbf{k}}$  be the center of  $I_{\mathbf{k}}$ . Since each side of the arc  $I_{\mathbf{k}}$  is  $\geq 3\delta(\mathcal{C})$ , the set  $\mathcal{C}$  has at least one element in the arc  $\{\mathbf{x} : |\mathbf{x} - \mathbf{z}_{\mathbf{k}}|_\infty \leq \delta(\mathcal{C})\} \subset I_{\mathbf{k}}$ . We form a subset  $\mathcal{C}'$  of  $\mathcal{C}$  by choosing exactly one such element for each  $\mathbf{k}$ . Then it is easy to see that  $\delta(\mathcal{C}') \leq 4\delta(\mathcal{C})$ , and  $\eta(\mathcal{C}') \geq \delta(\mathcal{C}) \geq (1/4)\delta(\mathcal{C}')$ . Thus,  $(1/4)\delta(\mathcal{C}') \leq \eta(\mathcal{C}') \leq 4\delta(\mathcal{C}')$ . In all of our discussion below, we will therefore assume that such a subset  $\mathcal{C}'$  has been

chosen. The rest of the elements of  $\mathcal{C}$  do not make any difference to the statements; e.g., we may just set the weights corresponding to the points not so chosen to be 0. Therefore, rather than complicating our notations, we will just identify  $\mathcal{C}'$  with  $\mathcal{C}$  in our notations. Thus, we assume that

$$(1/4)\delta(\mathcal{C}) \leq \eta(\mathcal{C}) \leq 4\delta(\mathcal{C}) \leq 4, \quad (3.15)$$

and that each  $I_{\mathbf{k},\mathcal{C}}$  contains exactly one element of  $\mathcal{C}$ . Then  $M \stackrel{\text{def}}{=} |\mathcal{C}| = m^q$ , and we will re-index  $\mathcal{C}$  by setting  $\mathbf{y}_{\mathbf{k}}$  to be the unique element of  $\mathcal{C} \cap I_{\mathbf{k}}$ .  $\square$

One of the important steps in the proof of Theorem 3.5 is the following lemma.

**Lemma 3.1** *Let  $m \geq 1$  be an integer, let  $\{I_{\mathbf{k}} : 0 \leq \mathbf{k} \leq m-1\}$  be a partition of  $\mathbb{T}^q$  as defined in (3.14),  $\mathcal{C} \stackrel{\text{def}}{=} \{\mathbf{y}_{\mathbf{k}}\}_{0 \leq \mathbf{k} \leq m-1} \subset \mathbb{T}^q$ , where each  $\mathbf{y}_{\mathbf{k}} \in I_{\mathbf{k}}$ , and let (3.15) be satisfied. Then there exists  $C > 0$  such that if  $\epsilon > 0$  and  $N = C\epsilon m$ , then we have for every  $P \in \mathbb{H}_N^q$ ,*

$$\left| \frac{1}{m^q} \sum_{0 \leq \mathbf{k} \leq m-1} |P(\mathbf{y}_{\mathbf{k}})| - \int_{\mathbb{T}^q} |P(\mathbf{z})| d\mu_q^*(\mathbf{z}) \right| \leq \epsilon \|P\|_1, \quad (3.16)$$

and

$$\left| \frac{1}{m^q} \sum_{0 \leq \mathbf{k} \leq m-1} P(\mathbf{y}_{\mathbf{k}}) - \int_{\mathbb{T}^q} P(\mathbf{z}) d\mu_q^*(\mathbf{z}) \right| \leq \epsilon \|P\|_1. \quad (3.17)$$

The lemma is proved in much greater generality in [20]. The ideas behind the proof are quite well known, e.g. [77, Section 4.9.1] or [67]. Here we give a simplified version of the proof in [20], using some ideas in [54, 64, 65]. Our objective is to highlight the use of localized kernels.

*Proof of Lemma 3.1* In this proof, let  $\delta = \delta(\mathcal{C})$ , and  $\eta = \eta(\mathcal{C})$ . We note that  $\delta(\mathcal{C}) \sim 1/m$ . Let  $N \geq 1$  be an integer to be chosen later, and  $P \in \mathbb{H}_N^q$ . Using the mean value theorem, it is easy to see that

$$\max_{\mathbf{z} \in I_{\mathbf{k}}} |P(\mathbf{x}) - P(\mathbf{y}_{\mathbf{k}})| \leq \frac{c}{m} \max_{1 \leq r \leq q} \|\partial_r P\|_{\infty}, \quad (3.18)$$

Consequently,

$$\begin{aligned} & \left| \frac{1}{m^q} \sum_{0 \leq \mathbf{k} \leq m-1} |P(\mathbf{y}_{\mathbf{k}})| - \int_{\mathbb{T}^q} |P(\mathbf{z})| d\mu_q^*(\mathbf{z}) \right| \\ &= \left| \sum_{0 \leq \mathbf{k} \leq m-1} \int_{I_{\mathbf{k}}} (|P(\mathbf{y}_{\mathbf{k}})| - |P(\mathbf{z})|) d\mu_q^*(\mathbf{z}) \right| \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{0 \leq \mathbf{k} \leq m-1} \int_{I_{\mathbf{k}}} |P(\mathbf{y}_{\mathbf{k}}) - P(\mathbf{z})| d\mu_q^*(\mathbf{z}) \\
&\leq \max_{1 \leq r \leq q} \frac{c}{m^q} \sum_{0 \leq \mathbf{k} \leq m-1} \|\partial_r P\|_{\infty, I_{\mathbf{k}}}.
\end{aligned} \tag{3.19}$$

Now, let  $S > q$  be an integer,  $h$  be an infinitely differentiable low pass filter,  $1 \leq r \leq q$ , and  $H_r(\mathbf{u}) = iu_r h(|\mathbf{u}|_2)$ . Then  $H_r$  is also infinitely differentiable. Hence, the fact that  $|\mathbf{t}|_2 \sim |\mathbf{t}|_{\infty}$  for all  $\mathbf{t} \in \mathbb{R}^q$  and the estimate (3.2) used with  $H_r$  in place of  $H$  shows that for  $\mathbf{t} \in \mathbb{T}^q$ ,

$$|\partial_r \Phi_N(h, \mathbf{t})| = N |\Phi_N(H_r, \mathbf{t})| \leq c \frac{N^{q+1}}{\max(1, (N|\mathbf{t}|_{\infty})^S)}. \tag{3.20}$$

We observe that for  $\mathbf{z} \in \mathbb{T}^q$ ,

$$\partial_r P(\mathbf{z}) = \int_{\mathbb{T}^q} P(\mathbf{t}) \partial_r \Phi_N(h, \mathbf{z} - \mathbf{t}) d\mu_q^*(\mathbf{t}),$$

and from this deduce that

$$\sum_{0 \leq \mathbf{k} \leq m-1} \|\partial_r P\|_{\infty, I_{\mathbf{k}}} \leq N \int_{\mathbb{T}^q} |P(\mathbf{t})| \left\{ \sum_{0 \leq \mathbf{k} \leq m-1} \max_{\mathbf{z} \in I_{\mathbf{k}}} |\Phi_N(H_r, \mathbf{z} - \mathbf{t})| \right\} d\mu_q^*(\mathbf{t}). \tag{3.21}$$

For the rest of the proof let  $R_r$  denote the maximum of the expression in the braces in the above formula for all  $\mathbf{t} \in \mathbb{T}^q$ . In view of translation invariance of the kernels  $\Phi_N$ , we may assume without loss of generality that

$$R_r = \sum_{0 \leq \mathbf{k} \leq m-1} \max_{\mathbf{z} \in I_{\mathbf{k}}} |\Phi_N(H_r, \mathbf{z} - (-\pi, \dots, -\pi))|. \tag{3.22}$$

We conclude using (3.19) that

$$\left| \frac{1}{m^q} \sum_{0 \leq \mathbf{k} \leq m-1} |P(\mathbf{y}_{\mathbf{k}})| - \int_{\mathbb{T}^q} |P(\mathbf{z})| d\mu_q^*(\mathbf{z}) \right| \leq c \frac{N}{m^{q+1}} \left( \max_{1 \leq r \leq q} R_r \right) \|P\|_1. \tag{3.23}$$

For the rest of the proof let  $\beta = (2\pi N)/m$ . For  $\ell = 0, 1, \dots, m-1$ , let

$$\mathcal{J}_{\ell} = \left\{ \mathbf{k} \in \mathbb{Z}^q : 1 \leq \mathbf{k} \leq m-1, \frac{2\pi \ell}{m} \leq \min_{\mathbf{z} \in I_{\mathbf{k}}} |\mathbf{z} + (\pi, \dots, \pi)|_{\infty} \leq \frac{2\pi(\ell+1)}{m} \right\},$$



and let  $n_\ell$  denote the number of elements in  $\mathcal{J}_\ell$ . Then  $n_\ell \sim \ell^{q-1}$ . Therefore, (3.20) shows that

$$\begin{aligned} R_r &\leq cN^q \sum_{\ell=0}^{m-1} \frac{1}{\max(1, (\beta\ell)^S)} n_\ell = cN^q \left\{ \sum_{\ell \leq \beta^{-1}} \ell^{q-1} + \beta^{-S} \sum_{\ell > \beta^{-1}} \ell^{q-1-S} \right\} \\ &\leq cN^q \beta^{-q} \leq cm^q. \end{aligned}$$

Moreover, the very last constant  $c$  above can be chosen independently of  $r$ . Hence, (3.23) shows that

$$\left| \frac{1}{m^q} \sum_{0 \leq \mathbf{k} \leq m-1} |P(\mathbf{y}_\mathbf{k})| - \int_{\mathbb{T}^q} |P(\mathbf{z})| d\mu_q^*(\mathbf{z}) \right| \leq c \frac{N}{m}.$$

Choosing  $N = \epsilon m/c$ , we arrive at (3.16).

Since

$$\begin{aligned} &\left| \frac{1}{m^q} \sum_{0 \leq \mathbf{k} \leq m-1} P(\mathbf{y}_\mathbf{k}) - \int_{\mathbb{T}^q} P(\mathbf{z}) d\mu_q^*(\mathbf{z}) \right| \\ &= \left| \sum_{0 \leq \mathbf{k} \leq m-1} \int_{I_\mathbf{k}} (P(\mathbf{y}_\mathbf{k}) - P(\mathbf{z})) d\mu_q^*(\mathbf{z}) \right| \\ &\leq \sum_{0 \leq \mathbf{k} \leq m-1} \int_{I_\mathbf{k}} |P(\mathbf{y}_\mathbf{k}) - P(\mathbf{z})| d\mu_q^*(\mathbf{z}) \\ &\leq \max_{1 \leq r \leq q} \frac{c}{m^{q+1}} \sum_{0 \leq \mathbf{k} \leq m-1} \|\partial_r P\|_{\infty, I_\mathbf{k}}, \end{aligned}$$

the same proof as above shows also that (3.17) holds as well.  $\square$

The other important ingredient in the proof of Theorem 3.5 is the following consequence of the Hahn–Banach theorem, known as the Krein extension theorem; see [21] for a recent proof. Let  $\mathbb{X}$  be a normed linear space,  $\mathcal{K}$  be a subset of its normed dual  $\mathbb{X}^*$ , and  $\mathcal{V}$  be a linear subspace of  $\mathbb{X}$ . We say that a linear functional  $x^* \in \mathcal{V}^*$  is positive on  $\mathcal{V}$  with respect to  $\mathcal{K}$  if  $x^*(f) \geq 0$  for every  $f \in \mathcal{V}$  with the property that  $y^*(f) \geq 0$  for every  $y^* \in \mathcal{K}$ .

**Theorem 3.6** *Let  $\mathbb{X}$  be a normed linear space, let  $\mathcal{K}$  be a bounded subset of its normed dual  $\mathbb{X}^*$ , let  $\mathcal{V}$  be a linear subspace of  $\mathbb{X}$ , and let  $x^* \in \mathcal{V}^*$  be positive on  $\mathcal{V}$  with respect to  $\mathcal{K}$ . We assume further that there exists  $v_0 \in \mathcal{V}$  such that  $\|v_0\|_{\mathbb{X}} = 1$  and*

$$\inf_{y^* \in \mathcal{K}} y^*(v_0) = \beta^{-1} > 0. \quad (3.24)$$

Then there exists an extension  $X^* \in \mathbb{X}^*$  of  $x^*$  which is positive on  $\mathbb{X}$  with respect to  $\mathcal{K}$  and satisfies

$$\|X^*\|_{\mathbb{X}^*} \leq \beta \sup_{y^* \in \mathcal{K}} \|y^*\|_{\mathbb{X}^*} x^*(v_0). \quad (3.25)$$

With this preparation, we are ready to prove Theorem 3.5.

*Proof of Theorem 3.5* As explained earlier, we may assume that  $M = m^q$  and  $\mathcal{C} = \{\mathbf{y}_\mathbf{k} : 0 \leq \mathbf{k} \leq m-1\}$ .

We take  $\epsilon = 1/4$  in Lemma 3.1 and conclude from (3.16) that

$$(3/4)\|P\|_1 \leq \frac{1}{m^q} \sum_{0 \leq \mathbf{k} \leq m-1} |P(\mathbf{y}_\mathbf{k})| \leq (5/4)\|P\|_1, \quad (3.26)$$

and from (3.17) that

$$\left| \frac{1}{m^q} \sum_{0 \leq \mathbf{k} \leq m-1} P(\mathbf{y}_\mathbf{k}) - \int_{\mathbb{T}^q} P(\mathbf{z}) d\mu_q^*(\mathbf{z}) \right| \leq (1/4)\|P\|_1 \leq \frac{1}{3m^q} \sum_{0 \leq \mathbf{k} \leq m-1} |P(\mathbf{y}_\mathbf{k})|. \quad (3.27)$$

Denote by  $\mathbb{X}$  the space  $\mathbb{R}^M$ , equipped with the norm

$$\|\mathbf{x}\| = \sum_{0 \leq \mathbf{k} \leq m-1} \mu_q^*(I_\mathbf{k}) |x_\mathbf{k}| \quad \text{where } \mathbf{x} = (x_\mathbf{k})_{0 \leq \mathbf{k} \leq m-1}.$$

For the set  $\mathcal{K}$ , we choose the set of coordinate functionals;  $z_\mathbf{k}^*(\mathbf{x}) = x_\mathbf{k}$ . Then  $\mathcal{K}$  is clearly a compact subset of  $\mathbb{X}^*$ . We consider the operator  $\mathcal{S} : \mathbb{H}_N^q \rightarrow \mathbb{R}^M$  given by  $P \mapsto (P(\mathbf{y}_\mathbf{k}))_{\mathbf{k}=0}^{m-1}$ , and take the subspace  $\mathcal{V}$  of  $\mathbb{X}$  to be the range of  $\mathcal{S}$ . The lower estimate in (3.26) shows that  $\mathcal{S}$  is invertible on  $\mathcal{V}$ . We define the functional  $x^*$  on  $\mathcal{V}$  by

$$x^*(\mathcal{S}(P)) = \int_{\mathbb{T}^q} P(\mathbf{z}) d\mu_q^*(\mathbf{z}) - \frac{1}{3m^q} \sum_{0 \leq \mathbf{k} \leq m-1} P(\mathbf{y}_\mathbf{k}), \quad P \in \mathbb{H}_N^q.$$

Moreover, if each  $P(\mathbf{y}_\mathbf{k}) \geq 0$ , then, using (3.27), we can conclude that

$$\int_{\mathbb{T}^q} P(\mathbf{z}) d\mu_q^*(\mathbf{z}) - \frac{1}{3m^q} \sum_{0 \leq \mathbf{k} \leq m-1} P(\mathbf{y}_\mathbf{k}) \geq \frac{1}{3m^q} \sum_{0 \leq \mathbf{k} \leq m-1} P(\mathbf{y}_\mathbf{k}) \geq 0.$$

Thus,  $x^*$  is positive on  $\mathcal{V}$  with respect to  $\mathcal{K}$ . The element  $(1, \dots, 1) \in \mathcal{V}$  serves as  $v_0$  in Theorem 3.6. Theorem 3.6 then implies that there exists a nonnegative functional

$X^*$  on  $\mathbb{R}^M$  that extends  $x^*$ . We may identify this functional with  $(\tilde{W}_{\mathbf{k}})_{\mathbf{k}=0}^{m-1} \in \mathbb{R}^M$  such that each  $\tilde{W}_{\mathbf{k}} \geq 0$ . The fact that  $X^*$  extends  $x^*$  means that for each  $P \in \mathbb{H}_N^q$ ,

$$\begin{aligned} \int_{\mathbb{T}^q} P(\mathbf{z}) d\mu_q^*(\mathbf{z}) &= \sum_{0 \leq \mathbf{k} \leq m-1} (\tilde{W}_{\mathbf{k}} + (1/(3m^q))P(\mathbf{y}_{\mathbf{k}})) \\ &= \sum_{0 \leq \mathbf{k} \leq m-1} w_{\mathbf{k}} P(\mathbf{y}_{\mathbf{k}}) \quad \text{where} \quad w_{\mathbf{k}} \stackrel{\text{def}}{=} \tilde{W}_{\mathbf{k}} + (1/(3m^q)). \end{aligned} \quad (3.28)$$

If we revert to the original set of points and set  $w_{\mathbf{k}} = 0$  if  $\mathbf{y}_{\mathbf{k}}$  is not in the subset chosen as in the remark following the statement of Theorem 3.5, then this is (3.12).

It was proved in [21, Theorem 5.8] that (3.12) implies (3.13).  $\square$

Next, we make a few comments about the numerical computation of the weights  $w_{\mathbf{k}}$ . We observe that the reduction of the original data set  $\{\mathbf{y}_{\mathbf{k}}\}$  may be done in several different ways and the weights  $w_{\mathbf{k}}$  are not uniquely defined either. Having made the reduction of the original set, a straightforward way to find the weights numerically is the following, see [34]. We minimize  $\sum_{0 \leq \mathbf{k} \leq m-1} w_{\mathbf{k}}^2$  subject to the conditions

$$\sum_{0 \leq \mathbf{k} \leq m-1} w_{\mathbf{k}} \exp(i\mathbf{j} \cdot \mathbf{y}_{\mathbf{k}}) = \begin{cases} 1, & \text{if } j = 0, \\ 0, & \text{otherwise,} \end{cases}$$

for  $0 \leq \mathbf{j} < N$ . This involves the solution of a linear system of equations whose matrix, the so-called *Gram matrix*, is given by

$$V_{\mathbf{j}, \ell} = \sum_{0 \leq \mathbf{k} \leq m-1} \exp(i(\mathbf{j} - \ell) \cdot \mathbf{y}_{\mathbf{k}}), \quad 0 \leq \mathbf{j}, \ell < N. \quad (3.29)$$

If  $(a_{\mathbf{j}})_{0 \leq \mathbf{j} < N}$  is an arbitrary vector and we take  $P(\mathbf{x}) = \sum_{0 \leq \mathbf{j} < N} a_{\mathbf{j}} \exp(i\mathbf{j} \cdot \mathbf{x})$ , then the Raleigh quotient for this matrix can be calculated to be

$$\frac{\sum_{0 \leq \mathbf{k} \leq m-1} |P(\mathbf{y}_{\mathbf{k}})|^2}{\sum_{0 \leq \mathbf{j} < N} |a_{\mathbf{j}}|^2} = \frac{\sum_{0 \leq \mathbf{k} \leq m-1} |P(\mathbf{y}_{\mathbf{k}})|^2}{\|P\|_2^2},$$

where we used the Parseval identity in the last step. Thus, the largest and smallest eigenvalues of  $V$ ,  $\lambda_{\max}$  and  $\lambda_{\min}$  are given by

$$\lambda_{\min} = \min_{P \in \mathbb{H}_N^q} \frac{\sum_{0 \leq \mathbf{k} \leq m-1} |P(\mathbf{y}_{\mathbf{k}})|^2}{\|P\|_2^2}, \quad \lambda_{\max} = \max_{P \in \mathbb{H}_N^q} \frac{\sum_{0 \leq \mathbf{k} \leq m-1} |P(\mathbf{y}_{\mathbf{k}})|^2}{\|P\|_2^2}, \quad (3.30)$$

see [29, Theorem 4.2.2, p. 176]. In practice, one has to choose  $N$  by trial and error so that the condition number of  $V$  is “reasonable”. We are tempted to solve a non-linear optimization problem with the additional requirement that the weights should be non-negative. It is our experience that the non-negativity of the weights is not important in practice, but an inequality of the form (3.13) is essential. For any given data, such

an inequality will, of course, hold with some constants depending on  $N$  and the data set. However, it is of interest to estimate the constants. It turns out that the constants involved are proportional to  $\lambda_{\max}$  and  $\lambda_{\min}$ .

We would like to point out another interesting fact. Suppose one wishes to find a least squares fit from  $\mathbb{H}_N^q$  to the data of the form  $\{(\mathbf{y}_{\mathbf{k}}, z_{\mathbf{k}})\}_{0 \leq \mathbf{k} \leq m-1}$ , that is, find  $a_{\mathbf{j}}$ 's to minimize

$$\sum_{0 \leq \mathbf{k} \leq m-1} |w_{\mathbf{k}}| \left| z_{\mathbf{k}} - \sum_{0 \leq \ell < N} a_{\ell} \exp(i\ell \cdot \mathbf{y}_{\mathbf{k}}) \right|^2$$

for a suitable choice of  $w_{\mathbf{k}}$ . This involves the solution of a linear system of equations where the matrix involved is  $G$ , defined by

$$G_{\mathbf{j}, \ell} = \sum_{0 \leq \mathbf{k} \leq m-1} |w_{\mathbf{k}}| \exp(i(\mathbf{j} - \ell) \cdot \mathbf{y}_{\mathbf{k}}).$$

As before, we have the bounds

$$\tilde{\lambda}_{\min} \sum_{0 \leq \mathbf{k} \leq m-1} |w_{\mathbf{k}}| |P(\mathbf{y}_{\mathbf{k}})|^2 \leq \|P\|_2^2 \leq \tilde{\lambda}_{\max} \sum_{0 \leq \mathbf{k} \leq m-1} |w_{\mathbf{k}}| |P(\mathbf{y}_{\mathbf{k}})|^2, \quad P \in \mathbb{H}_N^q, \quad (3.31)$$

where  $\tilde{\lambda}_{\min}$  and  $\tilde{\lambda}_{\max}$  are the smallest and largest eigenvalues of  $G$ .

To describe these results while keeping track of the constants on the various data sets and weights, etc., and also to simplify our notation in further theory, it is convenient to use a measure notation. The actual choice of the reduced data set and the weights plays no role in our theoretical consideration. Therefore, it is convenient to define a measure  $\nu$  that associates the mass  $w_{\mathbf{k}}$  with each  $\mathbf{y}_{\mathbf{k}}$ , that is, for any subset  $B \subset \mathbb{T}^q$ ,

$$\nu(B) = \sum_{\mathbf{k}: \mathbf{y}_{\mathbf{k}} \in B} w_{\mathbf{k}}.$$

We pause in our discussion to review some basic notions related to signed measures and introduce some notation before proceeding further. We recall that the total variation measure of any signed measure  $\mu$  is defined by

$$|\mu|(\mathcal{U}) \stackrel{\text{def}}{=} \sup \sum_{i=1}^{\infty} |\mu(U_i)|, \quad \mathcal{U} \subset \mathbb{T}^q,$$

where the supremum is taken over all countable partitions  $\{U_i\}$  into measurable sets of  $\mathcal{U}$ . For the measure  $\nu$  as defined above, one can easily deduce that  $|\nu|(B) = \sum_{\mathbf{k}: \mathbf{y}_{\mathbf{k}} \in B} |w_{\mathbf{k}}|$  for any subset  $B \subset \mathbb{T}^q$ . It is well known that any signed measure  $\mu$  on  $\mathbb{T}^q$  satisfies  $|\mu|(\mathbb{T}^q) < \infty$ . The support of a measure  $\mu$ , denoted by  $\text{supp}(\mu)$ , is the set of all  $\mathbf{x} \in \mathbb{T}^q$  such that for every open subset  $U$  of  $\mathbb{T}^q$  containing  $\mathbf{x}$ ,  $|\mu|(U) > 0$ .

If  $1 \leq p \leq \infty$ ,  $\mu$  is a (possibly signed) measure on  $\mathbb{T}^q$ ,  $B \subset \mathbb{T}^q$  is  $\mu$ -measurable, and  $f : B \rightarrow \mathbb{C}$  is  $\mu$ -measurable, then the  $L^p$  norm of  $f$  with respect to  $\mu$  is given by

$$\|f\|_{\mu;B} \stackrel{\text{def}}{=} \begin{cases} \left\{ \int_B |f(x)|^p d|\mu| \right\}^{1/p}, & \text{if } 1 \leq p < \infty, \\ |\mu| - \text{ess sup } |f(x)|_{x \in B}, & \text{if } p = \infty. \end{cases} \quad (3.32)$$

As before, we will omit the mention of  $B$  if  $B = \mathbb{T}^q$ , and to keep the notation consistent, will also omit the measure  $\mu$  from the notation if  $\mu = \mu_q^*$ . The space  $X^p(\mu)$  denotes the  $L^p(\mu)$  closure of the set of all trigonometric polynomials. In this notation, the estimate (3.13) becomes

$$\|P\|_{v;1} \sim \|P\|_1, \quad P \in \mathbb{H}_N^q,$$

and (3.31) can be written as

$$\tilde{\lambda}_{\min} \|P\|_2^2 \leq \|P\|_{v;2}^2 \leq \tilde{\lambda}_{\max} \|P\|_2^2.$$

In the sequel, we will not fix the measure  $\nu$  any more as above, and instead we use the notations  $\nu$ ,  $\mu$ , etc., to denote arbitrary measures. With a different measure  $\nu$ , the formulas (3.30) become

$$\lambda_{\min} \|P\|_2^2 \leq \|P\|_{\nu;2}^2 \leq \lambda_{\max} \|P\|_2^2, \quad P \in \mathbb{H}_n^q.$$

We are now ready to resume our discussion of the M–Z inequalities and related topics.

**Definition 3.3** (a) A (possibly signed) measure  $\mu$  is called a **quadrature measure** of order  $n$  when

$$\int_{\mathbb{T}^q} T d\mu = \int_{\mathbb{T}^q} T d\mu_q^*, \quad T \in \mathbb{H}_n^q. \quad (3.33)$$

(b) A (possibly signed) measure  $\mu$  is called a **Marcinkiewicz–Zygmund measure**, or M–Z measure, of order  $n$  when the following M–Z inequality

$$\int_{\mathbb{T}^q} |T| d|\mu| = \|T\|_{\mu;1} \leq c(n, \mu) \|T\|_1, \quad T \in \mathbb{H}_n^q, \quad (3.34)$$

is satisfied where  $c(n, \mu)$  is a constant independent of  $T$ . The smallest  $c$  that works in (3.34) will be denoted by  $\|\mu\|_n$ .

It can be shown easily that for each  $n \in \mathbb{N}$ ,  $\|\cdot\|_n$  is a norm on the space of Radon measures. Clearly, for every  $n \in \mathbb{N}$ ,  $\mu_q^*$  itself is an M–Z quadrature measure of order  $n$  with  $\|\mu_q^*\|_n = 1$ . In general, if  $\mu$  is an M–Z quadrature measure of order  $n$  for some  $n > 0$ , then  $\|\mu\|_n \geq 1$ .

In [21, Proposition 2.1, Theorem 5.4], we proved the following.

**Theorem 3.7** *Let  $\mu$  be a Radon measure and let  $n \geq 2$ . Then we have the following inequalities where the constants  $c$  are independent of both  $\mu$  and  $n$ .*

(a) *If  $1 \leq p < \infty$  then*

$$\|P\|_{\mu;p} \leq c \|\mu\|_n^{1/p} \|P\|_p, \quad P \in \mathbb{H}_n^q. \quad (3.35)$$

*Conversely, if for some  $p \in [1, \infty)$  and  $A = A(\mu, n) > 0$ ,*

$$\|P\|_{\mu;p} \leq cA^{1/p} \|P\|_p, \quad P \in \mathbb{H}_n^q,$$

*holds, then  $\|\mu\|_n \leq cA$ .*

(b) *For any  $r > 0$  and  $\mathbf{x} \in \mathbb{T}^q$  we have  $|\mu|(\{\mathbf{y} : \|\mathbf{x} - \mathbf{y}\|_\infty \leq r\}) \leq c\|\mu\|_n(r + 1/n)^q$ . Conversely, if there exists a constant  $A = A(\mu, n)$  such that for every  $r > 0$  and  $\mathbf{x} \in \mathbb{T}^q$ ,  $|\mu|(\{\mathbf{y} : \|\mathbf{x} - \mathbf{y}\|_\infty \leq r\}) \leq A(r + 1/n)^q$ , then  $A \leq c\|\mu\|_n$ .*

(c) *For any constant  $\alpha > 0$ , we have  $\|\mu\|_n \sim \|\mu\|_{\alpha n}$ , where the constants involved in the  $\sim$  relationship depend only on  $\alpha$  (and  $q$ ) but not on  $\mu$  or  $n$ .*

Once we compute the maximum eigenvalue of the Gram matrix  $V$  defined in (3.29), Theorem 3.7(a) allows us to estimate the constant involved in (3.13). Theorem 3.7(b) gives a geometric criteria for (3.13) without referring to trigonometric polynomials. Theorem 3.7(c) shows that if (3.13) holds for some  $n$ , then it holds also with equivalent constants for  $\alpha n$  for every  $\alpha > 1$  as well. In particular, even if a quadrature measure of order  $n$  supported on the minimal number of points,  $\dim(\mathbb{H}_{n/2}^q)$ , does not exist when  $q \geq 2$ , M-Z measures of order  $n$  with this property are plentiful.

### 3.4 Wavelet-like representation

Our starting point here is the following analogue of Theorem 3.1, where the novelty is that in the case of functions in  $X^\infty$ , their samples are used for approximation.

First, some notation. If  $\mu$  is a (possibly signed) measure and  $f \in L^1(\mu)$ , we define

$$\hat{f}(\mu; \mathbf{k}) = \int_{\mathbb{T}^q} f(\mathbf{t}) \exp(-i\mathbf{k} \cdot \mathbf{t}) d\mu(\mathbf{t}), \quad \mathbf{k} \in \mathbb{Z}^q.$$

If  $\mu = \mu_q^*$  then  $\hat{f}(\mu; \mathbf{k}) = \hat{f}(\mathbf{k})$ . If  $\mu$  is discretely supported measure then  $\hat{f}(\mu; \mathbf{k})$  is a discretized approximation to  $\hat{f}(\mathbf{k})$  as indicated by  $\mu$ . A common example is the discrete Fourier transform, obtained by letting  $\mu$  be supported on a set of dyadic points on  $\mathbb{T}^q$ . It is well known that the Fast Fourier Transform (FFT) algorithm allows a fast computation of the coefficients  $\hat{f}(\mu; \mathbf{k})$  for the appropriate measures  $\mu$ , and, for this reason, it is used widely in engineering applications. In the case when  $\mu$  is sparsely supported, there are new algorithms with sublinear complexity, see [28]. There are many other examples, including in particular, the so-called low discrepancy

quadrature rules, see, e.g., [13] for a detailed discussion and further references. Our interest here is in the case when  $\mu$  is an M–Z quadrature measure.

The analogue of the summability operator is defined by

$$\begin{aligned}\sigma_n(\mu; H, f, \mathbf{x}) &\stackrel{\text{def}}{=} \sum_{0 \leq \mathbf{k} < n} H\left(\frac{\mathbf{k}}{n}\right) \hat{f}(\mu; \mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x}) \\ &= \int_{\mathbb{T}^q} f(\mathbf{t}) \Phi_n(H; \mathbf{x} - \mathbf{t}) d\mu(\mathbf{t}), \quad n \in \mathbb{N}, \mathbf{x} \in \mathbb{T}^q, f \in L^1(\mu).\end{aligned}\tag{3.36}$$

We will adopt the same conventions as in Sect. 3.2 with respect to overloaded notations. For example, when  $H(\mathbf{t}) = h(|\mathbf{t}|_2)$ , we will write  $\sigma_n(\mu; h, f)$  in place of  $\sigma_n(\mu; H, f)$ .

The analogue of Theorem 3.1 is the following.

**Theorem 3.8** *Let  $n \in \mathbb{N}$ , let  $S > q$  be an integer,  $1 \leq p \leq \infty$ , and let  $\mu$  be an M–Z quadrature measure of order  $3n/2$ . Let  $h$  be an  $S$ -times continuously differentiable low pass filter.*

(a) *For for all  $P \in \mathbb{H}_{n/2}^q$ , we have  $\sigma_n(\mu; h, P) = P$ .*

(b) *We have for all  $f \in X^p(\mu)$ ,*

$$\|\sigma_n(\mu; h, f)\|_p \leq c \|\mu\|_n^{1-1/p} \|f\|_{\mu; p}.\tag{3.37}$$

*Consequently, if  $f \in X^\infty$ , then  $\|\sigma_n(\mu; h, f)\|_\infty \leq c \|\mu\|_n \|f\|_\infty$ , and, furthermore,*

$$E_{n,\infty}(f) \leq \|f - \sigma_n(\mu; h, f)\|_\infty \leq c \|\mu\|_n E_{n/2,\infty}(f).\tag{3.38}$$

(c) *If  $f \in L^1(\mu)$  is supported on a compact set  $K$  and  $V$  is an open set with  $K \subset V$ , then*

$$\|\sigma_n(\mu; h, f)\|_{\infty, \mathbb{T}^q \setminus V} \leq c \|f\|_{\mu; 1} n^{q-S},\tag{3.39}$$

*where, in addition to  $S$  and  $h$ , the constant  $c$  may depend on  $K$  and  $V$ .*

*Proof* If  $P \in \mathbb{H}_{n/2}^q$ , then for each  $\mathbf{x} \in \mathbb{T}^q$  we have  $P\Phi_n(h, \mathbf{x} - \circ) \in \mathbb{H}_{3n/2}^q$ . Since  $\mu$  is a quadrature measure of order  $3n/2$ ,

$$P(\mathbf{x}) = \int_{\mathbb{T}^q} P(\mathbf{t}) \Phi_n(h, \mathbf{x} - \mathbf{t}) d\mu_q^*(\mathbf{t}) = \int_{\mathbb{T}^q} P(\mathbf{t}) \Phi_n(h, \mathbf{x} - \mathbf{t}) d\mu(\mathbf{t}) = \sigma_n(\mu; h, P, \mathbf{x}).$$

This proves part (a). The proof of part (b) is almost verbatim the same as that of Theorem 2.1(b), except that multivariate notation needs to be used, and, in addition,

we need the fact that  $\mu$  is an M–Z measure of order  $3n/2$  to conclude that for each  $\mathbf{x} \in \mathbb{T}^q$

$$\int_{\mathbb{T}^q} |\Phi_n(h, \mathbf{x} - \mathbf{t})| d|\mu|(\mathbf{t}) \leq c \|\mu\|_n \int_{\mathbb{T}^q} |\Phi_n(h, \mathbf{x} - \mathbf{t})| d\mu_q^*(\mathbf{t}) \leq c \|\mu\|_n.$$

Part (c) is easy to prove using the localization estimate (3.2).  $\square$

As an immediate corollary, we note the following complement to the one-sided inequality (3.34), cf. Lemma 2.1 and (3.31).

**Corollary 3.2** *Let  $n \in \mathbb{N}$  and let  $\mu$  be an M–Z quadrature measure of order  $3n$ . Then for  $P \in \mathbb{H}_n^q$  and  $1 \leq p < \infty$ ,*

$$\|P\|_p \leq c \|\mu\|_n^{1-1/p} \|P\|_{\mu; p} \leq c_1 \|\mu\|_n \|P\|_p. \quad (3.40)$$

*Proof* We use (3.37) with  $P$  in place of  $f$  and  $\sigma_{2n}$  in place of  $\sigma_n$ . Since  $\sigma_{2n}(\mu; h, P) = P$ , this leads to the first inequality in (3.40). The second inequality is a consequence of Theorem 3.7(a).  $\square$

Next, we come to the definition and characterization of local smoothness classes. The local smoothness classes are defined exactly as in Definition 2.3, except that multivariate analogues of  $\mathbb{T}$ , “arc”, and smoothness classes are used. Our objective is to state the analogues of Theorems 2.5 and 2.6. Since frames in the sense of  $L^2$  cannot be defined using values of the target function at countably many points, and the topic of tight frames does not add anything new to our discussion, we will not delve into the topic of tight frames.

We say a sequence of measures  $\boldsymbol{\mu} = \{\mu_n\}$  has nested support when

$$m < n \implies \text{supp}(\mu_m) \subseteq \text{supp}(\mu_n).$$

In theoretical considerations for approximation based on values of the target function, we will usually assume that the data is available as a sequence  $\{\mathcal{C}_m\}$  of finite subsets of  $\mathbb{T}^q$ . By taking unions, we may assume without loss of generality that the sets are nested. The following proposition, proved in [49, Proposition 2.1], shows that the data reduction and construction of M–Z quadrature formulas can be done in a consistent manner.

**Proposition 3.1** *Let  $\{\mathcal{C}_m\}$  be a sequence of finite subsets of  $\mathbb{T}^q$  with  $\delta(\mathcal{C}_m) \sim 1/m$ , and let  $\mathcal{C}_m \subseteq \mathcal{C}_{m+1}$  for  $m \in \mathbb{N}$ . Then there exists a sequence of subsets  $\{\tilde{\mathcal{C}}_m \subseteq \mathcal{C}_m\}$ , where, for  $m \in \mathbb{N}$ , we have  $\delta(\tilde{\mathcal{C}}_m) \sim 1/m$ ,  $\tilde{\mathcal{C}}_m \subseteq \tilde{\mathcal{C}}_{m+1}$ , and  $\delta(\tilde{\mathcal{C}}_m) \leq 2\eta(\tilde{\mathcal{C}}_m)$ .*

**Definition 3.4** Let  $h : \mathbb{R} \rightarrow [0, \infty)$  be compactly supported and let  $\boldsymbol{\mu} = \{\mu_n\}_{n=0}^\infty$  be a sequence of Borel measures on  $\mathbb{T}^q$  with nested support. Let  $f \in \cap_{n=0}^\infty L^1(\mu_n)$ . Then, for each non-negative integer  $n$ , the band-pass operator  $\tau_n(\boldsymbol{\mu}; h, f)$  with respect to  $\boldsymbol{\mu}$



is given by

$$\begin{aligned}\tau_0(\boldsymbol{\mu}; h, f) &\stackrel{\text{def}}{=} \sigma_1(\mu_0; h, f) \text{ and } \tau_n(\boldsymbol{\mu}; h, f) \\ &\stackrel{\text{def}}{=} \sigma_{2^n}(\mu_n; h, f) - \sigma_{2^{n-1}}(\mu_{n-1}; h, f), \text{ for } n \in \mathbb{N}.\end{aligned}\quad (3.41)$$

To keep the notation consistent, we will generally omit the symbol  $\boldsymbol{\mu}$  if each  $\mu_n = \mu_q^*$ . If we must mention this measure for emphasis, then we will write  $\tau_j(\mu_q^*; h, f)$  in this case.

The kernels  $\Psi_j$  are defined in the same way as (2.27), except for obvious multivariate substitutions.

The analogue of Theorem 2.5 is the following statement.

**Theorem 3.9** *Let  $1 \leq p \leq \infty$ ,  $\gamma > 0$ ,  $f \in X^p$ , and let  $h$  be an infinitely differentiable low pass filter. Let  $\mathbf{v} = \{v_j\}_{j=0}^\infty$  be a sequence of measures such that each  $v_j$  is a quadrature measure of order  $2^{j+1}$ . Further, let  $\boldsymbol{\mu} = \{\mu_j\}_{j=0}^\infty$  be a sequence of measures such that each  $\mu_j$  is an  $M$ -Z quadrature measure of order  $3 \times 2^{j-1}$ .*

(a) *We have*

$$f = \sum_{j=0}^{\infty} \tau_j(h, f) = \sum_{j=0}^{\infty} \int_{\mathbb{T}^q} \tau_j(h, f, \mathbf{t}) \Psi_j(h, \circ - \mathbf{t}) dv_j(\mathbf{t})$$

*with convergence in the sense of  $X^p$ .*

(b) *If  $f \in L^2$  then*

$$\|f\|_2^2 \leq \sum_{j=0}^{\infty} \|\tau_j(h, f)\|_2^2 = \sum_{j=0}^{\infty} \int_{\mathbb{T}^q} |\tau_j(h, f, \mathbf{t})|^2 dv_j(\mathbf{t}) \leq 4\|f\|_2^2. \quad (3.42)$$

(c) *If  $f \in X^\infty$ , then*

$$f = \sum_{j=0}^{\infty} \tau_j(\boldsymbol{\mu}; h, f) = \sum_{j=0}^{\infty} \int_{\mathbb{T}^q} \tau_j(\mu_j; h, f, \mathbf{t}) \Psi_j(h, \circ - \mathbf{t}) dv_j(\mathbf{t}), \quad (3.43)$$

*where both the series converge uniformly.*

The interest in the above theorem is clearly when  $\mathbf{v}$  is a sequence of discretely supported measures. The proofs of this theorem and of the following analogue of Theorem 2.6 are verbatim the same as those of their univariate analogues, and therefore we omit them.

**Theorem 3.10** *Let  $1 \leq p \leq \infty$ ,  $\gamma > 0$ ,  $\mathbf{x}_0 \in \mathbb{T}^q$ ,  $f \in X^p$ , and let  $h$  be an infinitely differentiable low pass filter. Let  $\mu$  and  $\nu$  be sequences of measures as in Theorem 3.9. We assume further that  $\sup_{j \geq 0} \|\mu_j\|_{2j} < \infty$  and  $\sup_{j \geq 0} \|\nu_j\|_{2j} < \infty$ . Then the following are equivalent.*

- (a)  $f \in W_{\gamma,p}(\mathbf{x}_0)$ .
- (b) *There exists an arc  $I \ni \mathbf{x}_0$  such that*

$$\|\tau_j(h, f)\|_{\nu_j; p, I} = \mathcal{O}(2^{-j\gamma}). \quad (3.44)$$

- (c) *In the case when  $p = \infty$ ,*

$$\max_{\mathbf{t} \in \text{supp}(\nu_j) \cap I} |\tau_j(\mu; h, f, \mathbf{t})| = \mathcal{O}(2^{-j\gamma}). \quad (3.45)$$

*Remark* As before, the main interest is when  $\nu$  is a sequence of discretely supported measures with nested supports. We note that the polynomials  $\tau_j(h, f)$  in the estimates (3.44) are computed using the Fourier coefficients of  $f$ . The estimates (3.45) are more general in the sense that the sequence  $\mu$  can be a sequence of discrete measures as well, in which case, the values of  $f$  at the points in the supports of these measures are used. Of course, such a generalization is possible only if  $p = \infty$ , so that point evaluations are well defined.  $\square$

## 4 Periodic basis function (PBF) networks

### 4.1 Trigonometric polynomials and PBFs

We recall that a PBF network is a function of the form  $\mathbf{x} \mapsto \sum_{k=1}^N a_k G(\mathbf{x} - \mathbf{x}_k)$ ,  $\mathbf{x} \in \mathbb{T}^q$  where  $G \in X^\infty$  is called the *activation function*,  $N$  is the number of *neurons*, and  $\mathbf{x}_k \in \mathbb{T}^q$  are called *centers*. Thus, a PBF network is a translation network with activation function defined on  $\mathbb{T}^q$  and the centers in  $\mathbb{T}^q$  as well.

Starting with [51], we discovered a close connection between approximation by trigonometric polynomials and that by periodic basis function networks. While all the research known to us prior to [51] gave the degree of approximation by RBF networks in terms of a scaling parameter, the results in [51] appear to be the first of their kind where the theory was developed in terms of the number of neurons. Later on, it was observed by Schaback and Wendland [73], and independently in [42]—in the context of the so-called Gaussian networks—that it is the minimal separation among the centers, cf. Definition 3.2, of the network which leads to a complete theory of direct and converse theorems for approximation in this context. Accordingly, we will formulate our theorems here in terms of the minimal separation among the centers.

In the sequel, let  $G \in X^\infty(\mathbb{T}^q)$  be a fixed activation function satisfying the condition that  $\hat{G}(\mathbf{k}) \neq 0$  for any  $\mathbf{k} \in \mathbb{Z}^q$ . For  $f \in L^1$ , we may then define formally a “derivative” by the formula

$$\widehat{D_G(f)}(\mathbf{k}) \stackrel{\text{def}}{=} \frac{\hat{f}(\mathbf{k})}{\hat{G}(\mathbf{k})}, \quad \mathbf{k} \in \mathbb{Z}^q. \quad (4.1)$$

In the case when  $D_G(f) \in L^2$ , one says that  $f$  is in the native space of  $G$ . However, we will not need this concept. For our purpose, it is enough to note that if  $P \in \mathbb{H}_n^q$  for some  $n \in \mathbb{N}$ , then  $D_G P \in \mathbb{H}_n^q$  as well. In fact, we verify the following important observation:

**Proposition 4.1** *Let  $n \in \mathbb{N}$ ,  $P \in \mathbb{H}_n^q$ . Then*

$$P(\mathbf{x}) = \int_{\mathbb{T}^q} G(\mathbf{x} - \mathbf{y}) D_G(P, \mathbf{y}) d\mu_q^*(\mathbf{y}), \quad \mathbf{x} \in \mathbb{T}^q. \quad (4.2)$$

The proof of this proposition is a simple comparison of Fourier coefficients of both sides of (4.2). The basic idea in our proofs of the direct theorems as well as in wavelet-like representations is to discretize the integral in (4.2). The resulting estimate is shown in Theorem 4.1 below. In the remainder of this section, we use the notation

$$\mathfrak{m}_n = \mathfrak{m}_n(G) = \min_{|\mathbf{k}|_2 \leq n} |\hat{G}(\mathbf{k})|, \quad (4.3)$$

and

$$x_+ = \max(x, 0), \quad x \in \mathbb{R}.$$

**Theorem 4.1** *Let  $1 \leq p \leq \infty$ ,  $n \in \mathbb{N}$ ,  $N \in \mathbb{N}$ ,  $P \in \mathbb{H}_n^q$ , and let  $\nu$  be an M–Z quadrature measure of order  $n + N$ . Then*

$$\left\| P - \int_{\mathbb{T}^q} G(\circ - \mathbf{y}) D_G(P, \mathbf{y}) d\nu(\mathbf{y}) \right\|_{\infty} \leq c \| \nu \|_{n+N} \left( n^{\left( \frac{q}{p} - \frac{q}{2} \right)_+} \right) \frac{E_{N,\infty}(G)}{\mathfrak{m}_n} \| P \|_p. \quad (4.4)$$

We observe that in view of Theorem 3.5, an arbitrary finite subset with sufficiently high density content admits an M–Z quadrature measure of order  $n + N$  supported on this subset. In particular, such a measure  $\nu$  exists with  $|\text{supp}(\nu)| \sim (n + N)^q$ ,  $\| \nu \|_{n+N} \leq c$ , and  $\eta(\text{supp}(\nu)) \sim (n + N)^{-1}$ . Using such a choice of measure for  $\nu$ , the integral expression in (4.4) is a PBF network with  $\sim (n + N)^q$  neurons, the set of centers being  $\text{supp}(\nu)$ , and the minimal separation among its centers is  $\sim (n + N)^{-1}$ .

We note a corollary of Theorem 4.1.

**Corollary 4.1** *Let  $1 \leq p \leq \infty$ , and let  $G$  satisfy  $\hat{G}(\mathbf{k}) \neq 0$  for any  $\mathbf{k} \in \mathbb{Z}^q$ . Then the class of all translation networks with  $G$  as the activation function is dense in  $X^p$ .*

*Proof* If  $f \in X^p$  and  $\epsilon > 0$ , we find  $n \in \mathbb{N}$  and  $P \in \mathbb{H}_n^q$  such that  $\| f - P \|_p < \epsilon/2$ . For this  $n$ , we may find  $N$  large enough so that the network  $\mathcal{G}$  as constructed in (4.4) satisfies  $\| P - \mathcal{G} \|_p < \epsilon/2$ . Then  $\| f - \mathcal{G} \|_p < \epsilon$ .  $\square$

The estimate (4.4) is much more meaningful in the case when  $G$  is very smooth, that is, for every  $L > 0$ ,

$$\lim_{|\mathbf{k}|_2 \rightarrow \infty} |\mathbf{k}|_2^L |\hat{G}(\mathbf{k})| = 0.$$

We introduce the following definition.

**Definition 4.1** Let  $A > 0$ . We will say that  $G \in \mathcal{E}_A$  if each of the following conditions is satisfied:

1.  $G \in X^\infty$ ,
2. for all  $\mathbf{k} \in \mathbb{Z}^q$ ,  $\hat{G}(\mathbf{k}) \neq 0$ ,
- 3.

$$\limsup_{n \rightarrow \infty} \left( \frac{E_{An, \infty}(G)}{\mathfrak{m}_n} \right)^{1/n} < 1. \quad (4.5)$$

For  $G \in \mathcal{E}_A$ , Theorem 4.1 (used with  $N = An$ ) leads to the following corollary.

**Corollary 4.2** Let  $A > 0$  and  $G \in \mathcal{E}_A$ . Let  $1 \leq p \leq \infty$ ,  $n \in \mathbb{N}$ ,  $P \in \mathbb{H}_n^q$ , and let  $\nu$  be an  $M$ -Z quadrature measure of order  $(1 + A)n$ . Then there exists  $\rho = \rho(A, p)$  such that  $0 < \rho < 1$ , and

$$\left\| P - \int_{\mathbb{T}^q} G(\circ - \mathbf{y}) D_G(P, \mathbf{y}) d\nu(\mathbf{y}) \right\|_\infty \leq c \|\nu\|_n \rho^n \|P\|_p. \quad (4.6)$$

Many standard examples used in network expansions involve activation functions in  $\mathcal{E}_A$  for some  $A > 0$ , cf. [75] for the computation of Fourier transforms given in the examples below.

*Example 4.1* Periodization of the Gaussian.

$$G(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^q} \exp(-|\mathbf{x} - 2\pi\mathbf{k}|_2^2/2), \quad \hat{G}(\mathbf{k}) = (2\pi)^{q/2} \exp(-|\mathbf{k}|_2^2/2).$$

□

*Example 4.2* Periodization of the Hardy multiquadric.<sup>1</sup>

$$G(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^q} (\alpha^2 + |\mathbf{x} - 2\pi\mathbf{k}|_2^2)^{-1}, \quad \hat{G}(\mathbf{k}) = \frac{\pi^{(q+1)/2}}{\Gamma\left(\frac{q+1}{2}\right)\alpha} \exp(-\alpha|\mathbf{k}|_2).$$

□

<sup>1</sup> A Hardy multiquadric is a function of the form  $\mathbf{x} \mapsto (\alpha^2 + |\mathbf{x}|_2^2)^{-1}$ ,  $\mathbf{x} \in \mathbb{R}^q$ . It is one of the oft-used function in theory and applications of radial basis function networks. For a survey, see the paper [27] of Hardy.

**Example 4.3** Tensor product construction using the Poisson kernel. With  $0 < r < 1$ ,

$$G(\mathbf{x}) = \prod_{j=1}^q \frac{1-r^2}{1+r^2-2r\cos(x_j)}, \quad \hat{G}(\mathbf{k}) = r^{|\mathbf{k}|_1}.$$

□

For the proof of Theorem 4.1, we recall the following Nikolskii inequalities, see, e.g., [77, Section 4.9.4, p. 231].

**Lemma 4.1** *Let  $1 \leq p < r \leq \infty$ . Then for  $n \in \mathbb{N}$ ,*

$$\|P\|_r \leq cn^{\left(\frac{q}{p}-\frac{q}{r}\right)} \|P\|_p, \quad P \in \mathbb{H}_n^q. \quad (4.7)$$

We are now ready to prove Theorem 4.1. In the remainder of this section, we fix an infinitely differentiable low pass filter  $h$ .

*Proof of Theorem 4.1* In this proof, let  $R = \sigma_N(h, G)$ , so that  $R \in \mathbb{H}_N^q$ , and

$$\|G - R\|_\infty \leq cE_{N,\infty}(G). \quad (4.8)$$

Since  $\nu$  is a quadrature measure of order  $n + N$  and  $D_G(P) \in \mathbb{H}_n^q$ , we obtain

$$\begin{aligned} P(\mathbf{x}) &= \int_{\mathbb{T}^q} G(\mathbf{x} - \mathbf{y}) D_G(P, \mathbf{y}) d\mu_q^*(\mathbf{y}) \\ &= \int_{\mathbb{T}^q} R(\mathbf{x} - \mathbf{y}) D_G(P, \mathbf{y}) d\mu_q^*(\mathbf{y}) + \int_{\mathbb{T}^q} [G(\mathbf{x} - \mathbf{y}) - R(\mathbf{x} - \mathbf{y})] D_G(P, \mathbf{y}) d\mu_q^*(\mathbf{y}) \\ &= \int_{\mathbb{T}^q} R(\mathbf{x} - \mathbf{y}) D_G(P, \mathbf{y}) d\nu(\mathbf{y}) + \int_{\mathbb{T}^q} [G(\mathbf{x} - \mathbf{y}) - R(\mathbf{x} - \mathbf{y})] D_G(P, \mathbf{y}) d\mu_q^*(\mathbf{y}) \\ &= \int_{\mathbb{T}^q} G(\mathbf{x} - \mathbf{y}) D_G(P, \mathbf{y}) d\nu(\mathbf{y}) + \int_{\mathbb{T}^q} [G(\mathbf{x} - \mathbf{y}) - R(\mathbf{x} - \mathbf{y})] D_G(P, \mathbf{y}) d\mu_q^*(\mathbf{y}) \\ &\quad - \int_{\mathbb{T}^q} [G(\mathbf{x} - \mathbf{y}) - R(\mathbf{x} - \mathbf{y})] D_G(P, \mathbf{y}) d\nu(\mathbf{y}). \end{aligned}$$

Consequently,

$$\begin{aligned} &\left| P(\mathbf{x}) - \int_{\mathbb{T}^q} G(\mathbf{x} - \mathbf{y}) D_G(P, \mathbf{y}) d\nu(\mathbf{y}) \right| \\ &\leq \int_{\mathbb{T}^q} |G(\mathbf{x} - \mathbf{y}) - R(\mathbf{x} - \mathbf{y})| |D_G(P, \mathbf{y})| d\mu_q^*(\mathbf{y}) \\ &\quad + \int_{\mathbb{T}^q} |G(\mathbf{x} - \mathbf{y}) - R(\mathbf{x} - \mathbf{y})| |D_G(P, \mathbf{y})| d|\nu|(\mathbf{y}). \end{aligned}$$

Since  $\nu$  is an M-Z quadrature measure of order  $n + N$ ,  $\|\nu\|_{n+N} \geq 1$ , and we conclude that

$$\left\| P - \int_{\mathbb{T}^q} G(\circ - \mathbf{y}) D_G(P, \mathbf{y}) d\nu(\mathbf{y}) \right\|_{\infty} \leq c E_{N,\infty}(G) \{ \|D_G(P)\|_1 + \|D_G(P)\|_{\nu;1} \} \\ \leq c \|\nu\|_{n+N} E_{N,\infty}(G) \|D_G(P)\|_1. \quad (4.9)$$

Using Nikolskii inequalities (4.7) and recalling the definition (4.1) of  $\widehat{D_G(P)}(\mathbf{k})$ , we deduce that

$$\|D_G(P)\|_1^2 \leq \|D_G(P)\|_2^2 = \sum_{\mathbf{k}: |\mathbf{k}|_2 < n} \left| \frac{\hat{P}(\mathbf{k})}{\hat{G}(\mathbf{k})} \right|^2 \leq m_n^{-2} \sum_{\mathbf{k}: |\mathbf{k}|_2 < n} |\hat{P}(\mathbf{k})|^2 = m_n^{-2} \|P\|_2^2 \\ \leq c \left( n^{2\left(\frac{q}{p} - \frac{q}{2}\right)_+} \right) m_n^{-2} \|P\|_p^2.$$

The estimate (4.4) follows by substituting the above estimate into (4.9).  $\square$

While Theorem 4.1 shows that polynomials can be approximated well by PBF networks, the converse is also true. To describe this, we introduce some notation.

Let  $\{\mathbf{y}_j\}_{j=1}^M \subset \mathbb{T}^q$  and let  $n \in \mathbb{N}$  be an integer with

$$\min_{j \neq k} |\mathbf{y}_j - \mathbf{y}_k| \geq 1/n. \quad (4.10)$$

We note that this implies  $M \leq cn^q$ . In the sequel, we will assume tacitly that  $\{\mathbf{y}_j\}_{j=1}^M$  is one of the members of a sequence of finite subsets of  $\mathbb{T}^q$ . We assume that  $M$  and  $n$  are variables, and then the following constants are independent of these.

**Theorem 4.2** *Let  $\{\mathbf{y}_j\}_{j=1}^M \subset \mathbb{T}^q$ ,  $n \in \mathbb{N}$  be an integer satisfying (4.10). Let  $r \geq 0$ ,  $\mathbf{a} \in \mathbb{R}^M$ , and let*

$$\mathcal{G}(\mathbf{x}) = \sum_{j=1}^M a_j G(\mathbf{x} - \mathbf{y}_j), \quad \mathbf{x} \in \mathbb{T}^q.$$

*Then for  $N \in \mathbb{N}$ , we have*

$$\|(-\Delta)^{r/2} \mathcal{G} - (-\Delta)^{r/2} \sigma_N(h, \mathcal{G})\|_{\infty} \leq c_1 \left( n^{\left(\frac{q}{p} - \frac{q}{2}\right)_+} \right) \frac{E_{N,\infty}((-\Delta)^{r/2} G)}{m_{cn}((-\Delta)^{r/2} G)} \|(-\Delta)^{r/2} \mathcal{G}\|_p. \quad (4.11)$$

*In particular, if  $G \in \mathcal{E}_A$  for some  $A > 0$ , then there exists a  $\rho = \rho(A, p, r, G) \in (0, 1)$  such that*

$$\|(-\Delta)^{r/2} \mathcal{G} - (-\Delta)^{r/2} \sigma_{cAn}(h, \mathcal{G})\|_{\infty} \leq c_1 \rho^n \|(-\Delta)^{r/2} \mathcal{G}\|_p. \quad (4.12)$$

Before proceeding to the proof, we note the following corollary.

**Corollary 4.3** *Let  $A > 0$  and  $G \in \mathcal{E}_A$ . With notation as in Theorem 4.2, we have for  $n \in \mathbb{N}$ ,*

$$\|(-\Delta)^{r/2}\mathcal{G}\|_p \leq cn^r \|\mathcal{G}\|_p, \quad (4.13)$$

where  $c$  is a positive constant, independent of the choice of the coefficients or the nodes  $\{\mathbf{y}_j\}$ , as long as (4.10) holds.

*Proof* If  $n$  is large enough, (4.12) yields

$$\|(-\Delta)^{r/2}\mathcal{G} - (-\Delta)^{r/2}\sigma_{cAn}(h, \mathcal{G})\|_p \leq (1/2)\|(-\Delta)^{r/2}\mathcal{G}\|_p, \quad (4.14)$$

and since we may also use the same estimate with  $r = 0$ ,

$$\|\mathcal{G} - \sigma_{cAn}(h, \mathcal{G})\|_p \leq (1/2)\|\mathcal{G}\|_p. \quad (4.15)$$

This leads to

$$\|(-\Delta)^{r/2}\mathcal{G}\|_p \leq 2\|(-\Delta)^{r/2}\sigma_{cAn}(h, \mathcal{G})\|_p, \quad \|\sigma_{cAn}(h, \mathcal{G})\|_p \leq (3/2)\|\mathcal{G}\|_p. \quad (4.16)$$

Then using the Bernstein-type inequality in Theorem 3.2(b), we obtain

$$\|(-\Delta)^{r/2}\mathcal{G}\|_p \leq 2\|(-\Delta)^{r/2}\sigma_{cAn}(h, \mathcal{G})\|_p \leq c_1 n^r \|\sigma_{cAn}(h, \mathcal{G})\|_p \leq c_2 n^r \|\mathcal{G}\|_p.$$

If  $n$  is not large enough, then (4.13) is obtained by adjusting the constant factor.

The proof of Theorem 4.2 is yet another interesting application of the localization estimate (3.2). Using this estimate in a very critical manner, we proved the following in [5, Theorem 6.2].

**Proposition 4.2** *Let  $\{\mathbf{y}_j\}_{j=1}^M \subset \mathbb{T}^q$  and let  $n \in \mathbb{N}$  be an integer satisfying (4.10). Let  $1 \leq p \leq \infty$  and  $\mathbf{a} \in \mathbb{R}^M$ . Then we have*

$$c_2 n^{q/p'} |\mathbf{a}|_p \leq \left\| \sum_{j=1}^M a_j \Phi_m(h, \circ - \mathbf{y}_j) \right\|_p \leq c_3 n^{q/p'} |\mathbf{a}|_p, \quad (4.17)$$

for  $m \geq c_1 n$ .

*Proof of Theorem 4.2* We observe first that is enough to prove the theorem for  $r = 0$ . The general case follows by applying the result with the activation function  $(-\Delta)^{r/2}G$  in place of  $G$ . Let  $\mathbf{x} \in \mathbb{T}^q$ . We note that

$$\sigma_N(h, \mathcal{G}, \mathbf{x} - \mathbf{y}_j) = \sum_{k=1}^M a_j \sigma_N(h, G, \mathbf{x} - \mathbf{y}_j),$$

and  $M \leq cn^q$ , so that

$$\|\mathcal{G} - \sigma_N(h, \mathcal{G})\|_\infty \leq cE_{N,\infty}(G)|\mathbf{a}|_1 \leq cn^{q/2}E_{N,\infty}(G)|\mathbf{a}|_2. \quad (4.18)$$

To estimate  $|\mathbf{a}|_2$ , we fix  $m = Cn$  for a sufficiently large  $m$  so as to satisfy the condition in Proposition 4.2. Let  $\alpha$  be defined by

$$\alpha \stackrel{\text{def}}{=} \left( \frac{q}{p} - \frac{q}{2} \right)_+.$$

Then we can deduce with some calculation and with (4.17) applied with  $p = 2$  that

$$\begin{aligned} \|\mathcal{G}\|_p^2 &\geq c\|\sigma_m(h, \mathcal{G})\|_p^2 \geq cn^{-2\alpha}\|\sigma_m(h, \mathcal{G})\|_2^2 \\ &= cn^{-2\alpha} \sum_{j,\ell=1}^M a_j \overline{a_\ell} \sum_{\mathbf{k} \in \mathbb{Z}^q} h\left(\frac{|\mathbf{k}|_2}{m}\right)^2 |\hat{G}(\mathbf{k})|^2 \exp(-i\mathbf{k} \cdot (\mathbf{y}_j - \mathbf{y}_\ell)) \\ &= cn^{-2\alpha} \sum_{\mathbf{k} \in \mathbb{Z}^q} h\left(\frac{|\mathbf{k}|_2}{Cn}\right)^2 |\hat{G}(\mathbf{k})|^2 \left| \sum_{j=1}^M a_j \exp(-i\mathbf{k} \cdot \mathbf{y}_j) \right|^2 \\ &\geq cn^{-2\alpha} m_{Cn}(G)^2 \sum_{\mathbf{k} \in \mathbb{Z}^q} h\left(\frac{|\mathbf{k}|_2}{m}\right)^2 \left| \sum_{j=1}^M a_j \exp(-i\mathbf{k} \cdot \mathbf{y}_j) \right|^2 \\ &= cn^{-2\alpha} m_{Cn}(G)^2 \left\| \sum_{j=1}^M a_j \Phi_m(h, \circ - \mathbf{y}_j) \right\|_2^2 \\ &\geq cn^{-2\alpha+q} m_{Cn}(G)^2 |\mathbf{a}|_2^2. \end{aligned}$$

Thus,

$$n^{q/2}|\mathbf{a}|_2 \leq cn^\alpha m_{Cn}(G)^{-1} \|\mathcal{G}\|_p.$$

Finally, the estimate (4.11) is obtained by substituting the upper bound on  $|\mathbf{a}|_2$  from this estimate into (4.18).  $\square$

## 4.2 Direct and equivalence theorems

In this and the next subsections, we fix  $G \in \mathcal{E}_A$ . We will deal with two different sequences of measures; the sequence  $\nu$  will be a sequence of discrete measures, whose supports will give us the centers of the networks, and the sequence  $\mu$  will be the sequence of measures so that the information on the target function is given in terms of integrals with respect to the members of this sequence. It is a common practice in the literature on RBF networks to choose the centers the same as the points at which



the target function is sampled, but we wish to make it a point that this is not necessary. To avoid confusion, we make the following definition.

**Definition 4.2** Let  $A > 0$ , and  $\nu = \{\nu_j\}_{j=0}^\infty$  be a sequence of Borel measures. We say that  $\nu \in \mathfrak{M}_A$  if each of the following conditions is satisfied. Here, the constants may depend upon  $A$  and the sequence  $\nu$ , but not on  $j$  or the individual measures  $\nu_j$ .

1. Each  $\nu_j$  is a discrete measure, and the support  $C_j = \text{supp}(\nu_j)$  satisfies

$$\delta(C_j) \sim \eta(C_j) \sim 2^{-j}.$$

2. Each  $\nu_j$  is an M-Z quadrature measure of order  $(1 + A)2^j$ , and  $\|\nu\|_{(1+A)2^j} \leq c$ .
3. If  $j > m$  then  $C_m \subseteq C_j$ .

We point out an important example of measures in  $\mathfrak{M}_A$ . Let

$$ZD_j^q = \{\mathbf{k} \in \mathbb{Z}^q : -2^j + 1 \leq k_\ell \leq 2^j, \ell = 1, 2, \dots, q\}, \quad j = 0, 1, 2, \dots,$$

and

$$\mathbf{v}_{\mathbf{k},j} = \mathbf{k}\pi/2^j, \quad \mathbf{k} \in ZD_j^q, \quad j = 0, 1, 2, \dots \quad (4.19)$$

It can be verified using the corresponding univariate result repeatedly that the following quadrature formula holds.

$$\frac{1}{(2\pi)^q} \int_{\mathbb{T}^q} P(\mathbf{t}) d\mathbf{t} = \frac{1}{2^{q(j+1)}} \sum_{\mathbf{k} \in ZD_j^q} P(\mathbf{v}_{\mathbf{k},j}), \quad P \in \mathbb{H}_{2^{j+1}}^q. \quad (4.20)$$

Let  $m$  be the integer part of  $\log_2(1 + A)$ , and for  $j = 0, 1, \dots$ , let  $\nu_j^*$  be the measure  $\nu_j^*$  that associates the mass  $2^{-q(j+m+1)}$  with each  $\mathbf{v}_{\mathbf{k},j+m}$ ,  $\mathbf{k} \in ZD_{j+m}^q$ . Then the sequence  $\nu^* = \{\nu_j^*\}_{j=0}^\infty \in \mathfrak{M}_A$ . This fact can be checked using Lemma 2.1 for each coordinate.

Let  $\nu \in \mathfrak{M}_A$ . We define

$$\mathbb{G}_j = \mathbb{G}_j(\nu) \stackrel{\text{def}}{=} \{G(\circ - \mathbf{y}) : \mathbf{y} \in \text{supp}(\nu_j)\}, \quad j \in \mathbb{N}, \quad \mathbb{G}_0 = \{0\}, \quad (4.21)$$

and observe that, since  $\{\text{supp}(\nu_j)\}$  is a nested sequence,  $\mathbb{G}_j$  is a nested sequence of linear subspaces of  $X^\infty$  and that the closure of their union in the sense of  $L^p$  is  $X^p$ . For  $f \in L^p$ , we write

$$\text{dist}(L^p, f, \mathbb{G}_j) = \inf\{\|f - \mathcal{G}\|_p : \mathcal{G} \in \mathbb{G}_j\}.$$

In addition to the centers, we will assume in the remainder of this and the next subsection that  $\mu = \{\mu_j\}_{j=0}^\infty$  is a sequence of measures such that each  $\mu_j$  is an M-Z quadrature measure of order  $3 \times 2^{j-1}$ . The sequence  $\mu^*$  in which each  $\mu_j = \mu_q^*$  is

one such sequence. Our interest is both in this sequence and the case when the  $\mu_j$ 's are discretely supported.

The direct theorem can be stated as follows.

**Theorem 4.3** *Let  $A > 0$ ,  $G \in \mathcal{E}_A$ ,  $\mathbf{v} \in \mathfrak{M}_A$ . Then there exists  $\rho \in (0, 1)$  such that for  $1 \leq p \leq \infty$  and  $f \in L^p$  we have*

$$\begin{aligned} \text{dist}(L^p, f, \mathbb{G}_j(\mathbf{v})) &\leq \left\| f - \int_{\mathbb{T}^q} G(\circ - \mathbf{y}) D_G \left( \sigma_{2^j}(\mu_q^*; h, f), \mathbf{y} \right) dv_j(\mathbf{y}) \right\|_p \\ &\leq c\{E_{2^j, p}(f) + \rho^{2^j} \|f\|_p\}. \end{aligned} \quad (4.22)$$

Consequently, if  $\gamma > 0$  and  $f \in W_{\gamma, p}$ , then

$$\text{dist}(L^p, f, \mathbb{G}_j(\mathbf{v})) = \mathcal{O}(2^{-j\gamma}). \quad (4.23)$$

In the case when  $p = \infty$ , (4.22) remains valid provided that  $\sigma_{2^j}(\mu_q^*; h, f)$  is replaced by  $\sigma_{2^j}(\mu_j; h, f)$ , where  $\mu$  is a sequence as described earlier.

We note the following variations of (4.22).

**Corollary 4.4** *With the set up as in Theorem 4.3,*

$$\left\| f - \int_{\mathbb{T}^q} G(\circ - \mathbf{y}) D_G \left( \sigma_{2^j}(\mu_q^*; h, f), \mathbf{y} \right) dv_j^*(\mathbf{y}) \right\|_p \leq c\{E_{2^j, p}(f) + \rho^{2^j} \|f\|_p\}. \quad (4.24)$$

In the case when  $p = \infty$  and  $\mu \in \mathfrak{M}_{\max(A, 4)}$ , we also have

$$\left\| f - \int_{\mathbb{T}^q} G(\circ - \mathbf{y}) D_G \left( \sigma_{2^j}(\mu_j; h, f), \mathbf{y} \right) d\mu_j(\mathbf{y}) \right\|_{\infty} \leq c\{E_{2^j, \infty}(f) + \rho^{2^j} \|f\|_{\infty}\}. \quad (4.25)$$

We observe that the integral expression in (4.24) is a PBF network with centers at the dyadic points (4.19). This is in keeping with traditional results on PBF approximation where scaled integer translates are considered. However, unlike in the classical setting, we do not require any a priori conditions such as the Strang–Fix conditions to ensure any polynomial reproduction. In particular, constants are not included in the network. In the case when  $\mu_j$  is used in place of  $\mu_q^*$ , the information on  $f$  which is used in the construction of the network is the values of  $f$  at points in  $\text{supp}(\mu_j)$ , but we have different choices for the centers of the resulting network. A popular choice is to use points from the same sequence of data sets as centers as in (4.25). However, we may also use dyadic centers instead as in (4.24), potentially leading to faster computations.

All the constructions are linear operators on the information available on the target function. Finally, the bounds (4.22), (4.24), and (4.25), and also their variants where  $\sigma_{2j}(\mu_q^*; h, f)$  is replaced by  $\sigma_{2j}(\mu_j; h, f)$ , show that the PBF networks constructed there are bounded operators on the spaces involved. Therefore, the constructions are stable.

*Proof of Theorem 4.3* We need only to sketch the proof of (4.22); the other statements are immediate consequences. To prove (4.22), we use (4.6) with  $\sigma_{2j}(\mu_q^*; h, f)$  (or with  $\sigma_{2j}(\mu_j; h, f)$  respectively, when appropriate), recall that  $\|\nu_j\|_{2j} \sim \|\nu_j\|_{(1+A)2j} \leq c$ , and then use the resulting estimate together with the triangle inequality and (3.4) (or with (3.38) respectively) to arrive at (4.22).  $\square$

Converse theorems for approximation by elements of  $\mathbb{G}_j$  can be obtained using Corollary 4.3, cf. [12, Theorem 9.1, also Chapter 6.7]. We note only the analogue of the equivalence theorem without proof.

**Theorem 4.4** *Let  $1 \leq p \leq \infty$ ,  $\gamma > 0$ ,  $A > 0$ ,  $G \in \mathcal{E}_A$  and  $\nu \in \mathfrak{M}_A$ . Then the following are equivalent.*

- (a)  $f \in W_{\gamma, p}$ .
- (b)  $\text{dist}(L^p, f, \mathbb{G}_j(\nu)) = \mathcal{O}(2^{-j\gamma})$ .
- (c) We have

$$\left\| f - \int_{\mathbb{T}^q} G(\circ - \mathbf{y}) D_G \left( \sigma_{2j}(\mu_q^*; h, f), \mathbf{y} \right) d\nu_j^*(\mathbf{y}) \right\|_p = \mathcal{O}(2^{-j\gamma}). \quad (4.26)$$

- (d) If  $p = \infty$ , then each of the above statements is also equivalent to

$$\left\| f - \int_{\mathbb{T}^q} G(\circ - \mathbf{y}) D_G \left( \sigma_{2j}(\mu_j; h, f), \mathbf{y} \right) d\nu_j(\mathbf{y}) \right\|_{\infty} = \mathcal{O}(2^{-j\gamma}).$$

We remark that the statement (a) in the above theorem is independent of the sequence  $\nu$ . Therefore, the part (a) above implies part (b) for every  $\nu \in \mathfrak{M}_A$ . On the other hand, if part (b) holds for some  $\nu \in \mathfrak{M}_A$ , then part (a) holds. In particular, when approximating a function from  $W_{\gamma, p}$ , the asymptotic degree of approximation by PBF networks does not depend upon the choice of centers and weights—encoded in the measures  $\nu$ —as long as  $\nu \in \mathfrak{M}_A$ .

#### 4.3 Wavelet-like representation using PBFs

The wavelet-like representations using PBF networks with activation function in  $\mathcal{E}_A$  for some  $A > 0$  are very similar to those for multivariate trigonometric polynomials. This can be shown by using Theorem 4.1. We summarize them below. We will sketch the proofs only, except for the proof of the frame property which requires some new ideas.

In the sequel, let  $A > 0$  and let  $G \in \mathcal{E}_A$  be fixed. Also, let  $h$  be a fixed, infinitely differentiable low pass filter. We will also make the same assumptions for the sequence of measures  $\mathbf{v} = \{\nu_j\}_{j=0}^\infty$  and  $\boldsymbol{\mu} = \{\mu_j\}_{j=0}^\infty$ ; namely,  $\mathbf{v} \in \mathfrak{M}_A$  and each  $\mu_j$  is an M-Z quadrature measure of order  $3 \times 2^{j-1}$ .

First, we consider the operators

$$\mathcal{S}_j^*(f, \mathbf{x}) \stackrel{\text{def}}{=} \int_{\mathbb{T}^q} G(\mathbf{x} - \mathbf{y}) D_G(\sigma_{2^j}(\mu_q^*; h, f), \mathbf{y}) d\nu_j(\mathbf{y}), \quad j = 0, 1, \dots, f \in L^p, p \in [1, \infty], \quad (4.27)$$

and the corresponding frame operators

$$\mathcal{T}_j^*(f, \mathbf{x}) = \begin{cases} \mathcal{S}_0^*(f, \mathbf{x}), & \text{if } j = 0, \\ \mathcal{S}_j^*(f, \mathbf{x}) - \mathcal{S}_{j-1}^*(f, \mathbf{x}), & \text{if } j \in \mathbb{N}. \end{cases} \quad (4.28)$$

The variants when the information about  $f$  is in terms of the sequence  $\boldsymbol{\mu}$  are given by

$$\mathcal{S}_j(f, \mathbf{x}) \stackrel{\text{def}}{=} \int_{\mathbb{T}^q} G(\mathbf{x} - \mathbf{y}) D_G(\sigma_{2^j}(\mu_j; h, f), \mathbf{y}) d\nu_j(\mathbf{y}), \quad j = 0, 1, \dots, f \in X^\infty, \quad (4.29)$$

and

$$\mathcal{T}_j(f, \mathbf{x}) = \begin{cases} \mathcal{S}_0(f, \mathbf{x}), & \text{if } j = 0, \\ \mathcal{S}_j(f, \mathbf{x}) - \mathcal{S}_{j-1}(f, \mathbf{x}), & \text{if } j \in \mathbb{N}. \end{cases} \quad (4.30)$$

Theorem 4.1 leads immediately to the following estimates for these operators.

**Proposition 4.3** *There exists  $\rho \in (0, 1)$  such that, if  $1 \leq p \leq \infty$  and  $f \in X^p$ , then we have*

$$\|\mathcal{S}_j^*(f) - \sigma_{2^j}(\mu_q^*; h, f)\|_p \leq c\rho^{2^j} \|\sigma_{2^j}(\mu_q^*; h, f)\|_p \leq c\rho^{2^j} \|f\|_p. \quad (4.31)$$

For any measure  $\tilde{\nu}$  on  $\mathbb{T}^q$ ,

$$\|\mathcal{T}_j^*(f) - \tau_j(\mu_q^*; h, f)\|_{\tilde{\nu}; p} \leq c\rho^{2^j} \|f\|_p. \quad (4.32)$$

Analogous statements hold if  $p = \infty$  and  $\mathcal{S}_j^*(f)$  (respectively,  $\mathcal{T}_j^*(f)$ ,  $\sigma_{2^j}(\mu_q^*; h, f)$ ,  $\tau_j(\mu_q^*; h, f)$ ) is replaced by  $\mathcal{S}_j(f)$  (respectively,  $\mathcal{T}_j(f)$ ,  $\sigma_{2^j}(\mu_j; h, f)$ ,  $\tau_j(\boldsymbol{\mu}; h, f)$ ).

The following two theorems, analogous to Theorems 4.5 and 3.10, are immediate consequences of these theorems and the above proposition, except for part (b) of Theorem 4.5 below.

**Theorem 4.5** *Let  $1 \leq p \leq \infty$ ,  $\gamma > 0$ ,  $f \in X^p$ , and let  $G$  and the measures  $\mathbf{v}$ ,  $\boldsymbol{\mu}$  be as described earlier.*

(a) We have

$$f = \sum_{j=0}^{\infty} \mathcal{T}_j^*(f) \quad (4.33)$$

with convergence in the sense of  $X^p$ .

(b) If  $f \in L^2$  then

$$\|f\|_2^2 \sim \sum_{j=0}^{\infty} \|\mathcal{T}_j^*(f)\|_2^2, \quad (4.34)$$

where the constants are independent of  $f$ , but may depend upon  $G$ .

(c) If  $f \in X^\infty$ , then we have the uniformly convergent expansion

$$f = \sum_{j=0}^{\infty} \mathcal{T}_j(f) \quad (4.35)$$

The analogue of Theorem 3.10 is the following:

**Theorem 4.6** Let  $1 \leq p \leq \infty$ ,  $\gamma > 0$ ,  $\mathbf{x}_0 \in \mathbb{T}^q$ ,  $f \in X^p$ , and let  $G$  and the measures  $\nu, \mu$  be as described earlier. Then the following are equivalent.

(a)  $f \in W_{\gamma,p}(\mathbf{x}_0)$ .

(b) There exists an arc  $I \ni \mathbf{x}_0$  such that

$$\|\mathcal{T}_j^*(f)\|_{\nu_j;p,I} = \mathcal{O}(2^{-j\gamma}). \quad (4.36)$$

(c) In the case when  $p = \infty$ , one may replace (4.36) by

$$\max_{\mathbf{t} \in \text{supp}(\nu_j) \cap I} |\mathcal{T}_j(f, \mathbf{t})| = \mathcal{O}(2^{-j\gamma}).$$

Part (b) of Theorem 4.5 is not immediately obvious. In order to prove this part, we first state a lemma.

**Lemma 4.2** Let  $f \in L^2$ , and let  $\rho$  be as in Proposition 4.3. Then for  $N \in \mathbb{N}$  and  $j \in \mathbb{N}$ ,

$$(1/4)\|f\|_2^2 \leq \|\sigma_{2N}(\mu_q^*; h, f)\|_2^2 + \sum_{j=N+1}^{\infty} \|\tau_j(\mu_q^*; h, f)\|_2^2 \leq \|f\|_2^2, \quad (4.37)$$

and

$$\left| \|\mathcal{T}_j^*(f)\|_2^2 - \|\tau_j(\mu_q^*; h, f)\|_2^2 \right| \leq c\rho^{2j} \|f\|_2^2. \quad (4.38)$$

*Proof* Using the same argument as in the proof of (2.36) in the proof of Theorem 2.5, we deduce that for all  $\mathbf{k} \in \mathbb{Z}^q$

$$1 \leq 2 \left( h \left( \frac{|\mathbf{k}|_2}{N} \right) \right)^2 + 4 \sum_{j=N+1}^{\infty} \left( h \left( \frac{|\mathbf{k}|_2}{2^j} \right) - h \left( \frac{|\mathbf{k}|_2}{2^{j-1}} \right) \right)^2 \leq 4.$$

Then, again as in the proof of Theorem 2.5, we use the Parseval identity to obtain (4.37).

Next, we use (4.32) with  $v = \mu_q^*$ ,  $A = \mathbb{T}^q$ , and  $p = 2$  to observe that

$$\begin{aligned} \left| \|\mathcal{T}_j^*(f)\|_2^2 - \|\tau_j(\mu_q^*; h, f)\|_2^2 \right| &= (\|\mathcal{T}_j^*(f)\|_2 + \|\tau_j(\mu_q^*; h, f)\|_2)(\|\mathcal{T}_j^*(f)\|_2 - \|\tau_j(\mu_q^*; h, f)\|_2) \\ &\leq c\|f\|_2 \|\mathcal{T}_j^*(f) - \tau_j(\mu_q^*; h, f)\|_2 \leq c\rho^{2j} \|f\|_2^2. \end{aligned}$$

This is precisely (4.38).  $\square$

*Proof of Theorem 4.5* (b). Let  $N \in \mathbb{N}$  be fixed but to be chosen later. In light of (4.38),

$$\left| \sum_{j=N+1}^{\infty} \|\mathcal{T}_j^*(f)\|_2^2 - \sum_{j=N+1}^{\infty} \|\tau_j(\mu_q^*; h, f)\|_2^2 \right| \leq c\|f\|_2^2 \sum_{j=N+1}^{\infty} \rho^{2j}. \quad (4.39)$$

Also, (4.31) implies that

$$\left| \|\mathcal{S}_N^*(f)\|_2 - \|\sigma_{2^N}(\mu_q^*; h, f)\|_2 \right| \leq c_1 \rho^{2^N} \|\sigma_{2^N}(\mu_q^*; h, f)\|_2. \quad (4.40)$$

We now choose  $N$  such that

$$c \sum_{j=N+1}^{\infty} \rho^{2^j} < 1/8, \quad c_1 \rho^{2^N} < 1/2, \quad (4.41)$$

where  $c$  and  $c_1$  are as in the previous two displayed estimates. Then

$$\|\mathcal{S}_N^*(f)\|_2 \leq (3/2) \|\sigma_{2^N}(\mu_q^*; h, f)\|_2 \leq 3 \|\mathcal{S}_N^*(f)\|_2. \quad (4.42)$$

With this preparation, we first prove the upper bound on  $\|f\|_2^2$ . The first inequality in (4.37) yields

$$(1/4) \|f\|_2^2 \leq 4 \|\mathcal{S}_N^*(f)\|_2^2 + \sum_{j=N+1}^{\infty} \|\mathcal{T}_j^*(f)\|_2^2 + (1/8) \|f\|_2^2,$$

that is,

$$(1/32) \|f\|_2^2 \leq \|\mathcal{S}_N^*(f)\|_2^2 + \sum_{j=N+1}^{\infty} \|\mathcal{T}_j^*(f)\|_2^2. \quad (4.43)$$

Since

$$\|\mathcal{S}_N^*(f)\|_2^2 \leq \left( \sum_{j=0}^N \|\mathcal{T}_j^*(f)\|_2 \right)^2 \leq (N+1) \sum_{j=0}^N \|\mathcal{T}_j^*(f)\|_2^2,$$

(4.43) shows that

$$c\|f\|_2^2 \leq \sum_{j=0}^{\infty} \|\mathcal{T}_j^*(f)\|_2^2. \quad (4.44)$$

The lower bound is easier. For each  $j \geq 0$ , we have  $\|\mathcal{T}_j^*(f)\|_2 \leq c\|f\|_2$ . Therefore, using (4.37) and (4.39) and keeping in mind our choice of  $N$  as in (4.41), we can conclude that

$$\begin{aligned} \sum_{j=0}^{\infty} \|\mathcal{T}_j^*(f)\|_2^2 &= \sum_{j=0}^N \|\mathcal{T}_j^*(f)\|_2^2 + \sum_{j=N+1}^{\infty} \|\mathcal{T}_j^*(f)\|_2^2 \\ &\leq c(N+1)\|f\|_2^2 + \sum_{j=N+1}^{\infty} \|\tau_j(\mu_q^*; h, f)\|_2^2 + (1/8)\|f\|_2^2 \\ &\leq (c(N+1) + 1 + 1/8)\|f\|_2^2. \end{aligned}$$

This completes the proof.  $\square$

We note in closing that with a proper normalization, we may assume that  $\hat{G}(0) = 1$ . Then  $\mathfrak{m}_n(G) \geq 1$  for all  $n$ . Hence, the statement that  $G \in \mathcal{E}_A$  implies, in particular, that

$$\limsup_{m \rightarrow \infty} E_{m,\infty}(G)^{1/m} < 1.$$

This, in turn, implies that  $G$  is analytic on  $\mathbb{T}^q$ . There are many examples of activation functions which are used in practice, notably the periodization of the Wendland functions, or Green's functions of the operators  $(-\Delta)^{r/2}$ , for which this condition is not satisfied. The analogues of Theorems 4.1 and 4.2 are then much weaker; they are given in [49] in a very general context. Direct and converse theorems for approximation with networks with such activation functions are also obtained in that paper. However, the estimates there are not strong enough to obtain the characterization of local smoothness classes.

It is worthwhile to comment about the relationship of our results with those in the papers [14] by D  ng and Micchelli and [40] of Maiorov.

- The paper [40] deals with approximation in  $L^2$ , and the paper [14] deals with  $L^p$ ,  $1 < p < \infty$ . Our paper includes both  $p = 1$  and the case of continuous functions when  $p = \infty$ .

- The paper [14] deals with Korobov spaces rather than Sobolev spaces in the sense of our definitions. While our methods can be extended to the case of Korobov spaces, the papers in their present form are not comparable.
- The papers [14,40] give the bounds on approximation in terms of the number of neurons. Our paper gives the bounds in terms of the minimal separation among centers. In the case of uniform grids, the two concepts coincide. In this case, the upper bound in [40] is similar to ours, and the ideas behind its proof are essentially similar to those used in this paper. Both of these ideas are originally developed in [51].
- The papers [14,40] give lower bounds in the sense of *worst case complexity*; i.e., lower distance bounds. Our focus is on individual functions. For example, the lower bound in Theorem 1.1 of [40] implies that there **exists some**  $f$  in the function class under consideration for which the lower bound applies. The converse theorems in this paper are conceptually quite different. They state that for each **individual function**  $f$ , without any prior knowledge of the class to which it belongs, the rate of decrease of the degree of approximation *implies* the smoothness class to which it belongs.
- The paper [14] depends upon dyadic decompositions as is customary in the study of Korobov spaces. Preliminary numerical experiments suggest that the hyperbolic cross versions of the operators considered in [14] (or our operators for that matter) are not localized. We use the term wavelet-like representation to mean that the coefficients characterize local smoothness classes analogous to classical wavelet expansions. In this sense, it is an open problem to obtain a wavelet-like representation that characterizes local Korobov spaces.
- There is an impressive lower bound in [14] which suggests that the best activation function in the case of approximation of Korobov spaces is the Korobov kernel itself. Intuitively, this deep result is somewhat expected, since the Korobov spaces are, in a limiting sense, translation networks with the Korobov kernel as the activation function. Analogous results for approximation of periodic functions are given in [52,53]. The current paper does not deal with the question of the choice of an optimal activation function.

## 5 Further extensions

### 5.1 Jacobi expansions

Given  $\alpha > -1$  and  $\beta > -1$ , the Jacobi weights are defined by

$$w_{\alpha,\beta}(x) \stackrel{\text{def}}{=} \begin{cases} (1-x)^\alpha(1+x)^\beta, & \text{if } -1 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

For  $1 \leq p < \infty$ , the space  $L^p(\alpha, \beta)$  is defined as the space of (equivalence classes of) functions  $f$  with



$$\|f\|_{\alpha, \beta; p} \stackrel{\text{def}}{=} \left( \int_{-1}^1 |f(x)|^p w_{\alpha, \beta}(x) dx \right)^{1/p} < \infty.$$

The symbol  $X^p(\alpha, \beta)$  denotes  $L^p(\alpha, \beta)$ , if  $1 \leq p < \infty$ , and  $C([-1, 1])$ , the space of continuous functions on  $[-1, 1]$  with the maximum norm  $\|\circ\|_\infty$ , if  $p = \infty$ . The space of all algebraic polynomials of degree at most  $n$  will be denoted by  $\Pi_n$ .

There exists a unique system of orthonormalized Jacobi polynomials  $\{p_j^{(\alpha, \beta)}(x) = \gamma_j(\alpha, \beta)x^j + \dots\}$ ,  $\gamma_j(\alpha, \beta) > 0$  such that for integer  $j, \ell = 0, 1, \dots$ ,

$$\int_{-1}^1 p_j^{(\alpha, \beta)}(x) p_\ell^{(\alpha, \beta)}(x) w_{\alpha, \beta}(x) dx = \begin{cases} 1, & \text{if } j = \ell, \\ 0, & \text{otherwise.} \end{cases}$$

The uniqueness of the system implies that  $p_j^{(\beta, \alpha)}(x) = (-1)^j p_j^{(\alpha, \beta)}(-x)$ ,  $x \in \mathbb{R}$ ,  $j = 0, 1, \dots$ . Therefore, we may assume in the sequel that  $\alpha \geq \beta$ . We will assume also that  $\alpha \geq \beta \geq -1/2$ .

If  $f \in L^1(\alpha, \beta)$ , then, in this subsection, we define the Jacobi coefficients by

$$\hat{f}(j) \stackrel{\text{def}}{=} \hat{f}(\alpha, \beta; j) \stackrel{\text{def}}{=} \int_{-1}^1 f(y) p_j^{(\alpha, \beta)}(y) w_{\alpha, \beta}(y) dy, \quad j = 0, 1, \dots,$$

so that the formal Jacobi expansion of  $f$  is given by  $\sum_{j \geq 0} \hat{f}(j) p_j^{(\alpha, \beta)}$ . We define the summability operator analogous to  $\sigma_n$  as follows. Let  $h: [0, \infty) \rightarrow \mathbb{R}$  be a compactly supported function. We define

$$\Phi_n(\alpha, \beta; h, x, y) = \sum_{j=0}^{\infty} h\left(\frac{j}{n}\right) p_j^{(\alpha, \beta)}(x) p_j^{(\alpha, \beta)}(y),$$

and, for  $f \in L^1(\alpha, \beta)$ ,

$$\begin{aligned} \sigma_n(\alpha, \beta; h, f, x) &= \sum_{j=0}^{\infty} h\left(\frac{j}{n}\right) \hat{f}(j) p_j^{(\alpha, \beta)}(x) \\ &= \int_{-1}^1 f(y) \Phi_n(\alpha, \beta; h, x, y) w_{\alpha, \beta}(y) dy, \quad n \in \mathbb{N}. \end{aligned}$$

Using a result on the Cesàro means of the Jacobi expansion, see below, it is fairly easy to show that these operators are uniformly bounded. To describe this, we first recall the definition of Cesàro means. If  $\kappa > -1$ , the Cesàro means of order  $\kappa$  of

$f \in L^1(\alpha, \beta)$  are defined by

$$C_n^{[\kappa]}(\alpha, \beta; f, x) \stackrel{\text{def}}{=} \binom{n+\kappa}{\kappa}^{-1} \sum_{j=0}^n \binom{n-j+\kappa}{\kappa} \hat{f}(j) p_j^{(\alpha, \beta)}(x) \quad (5.1)$$

The following theorem is well known, see, e.g., [1, 76].

**Theorem 5.1** *Let  $\alpha, \beta \geq -1/2$ ,  $\kappa > \max(\alpha, \beta) + 1/2$  be an integer,  $1 \leq p \leq \infty$ , and  $f \in X^p(\alpha, \beta)$ . Then*

$$\|C_n^{[\kappa]}(\alpha, \beta; f)\|_{\alpha, \beta; p} \leq c \|f\|_{\alpha, \beta; p}, \quad (5.2)$$

for  $n \in \mathbb{N}$ .

Using a summation by parts argument as done in [26, Theorem 71, p. 128], Theorem 5.1 (used with  $S$  in place of  $\kappa$ ) leads immediately to the following corollary.

**Corollary 5.1** *Let  $S > \max(\alpha, \beta) + 1/2$ . If  $\{h_j\}$  is a sequence of real numbers, so that  $h_j \rightarrow 0$  as  $j \rightarrow \infty$ , and*

$$\sum_{j=0}^{\infty} (j+1)^S |\Delta^{S+1} h_j| < \infty, \quad (5.3)$$

where  $\Delta$  here is the forward difference operator, then for  $1 \leq p \leq \infty$  and  $f \in X^p(\alpha, \beta)$ ,

$$\begin{aligned} \left\| \sum_{j=0}^{\infty} h_j \hat{f}(j) p_j^{(\alpha, \beta)} \right\|_p &= \left\| \sum_{j=0}^{\infty} \binom{j+S}{S} C_j^{[S]}(\alpha, \beta; f) \Delta^{S+1} h_j \right\|_p \\ &\leq c \left( \sum_{j=0}^{\infty} (j+1)^S |\Delta^{S+1} h_j| \right) \|f\|_{\alpha, \beta; p}. \end{aligned} \quad (5.4)$$

In particular, if  $h$  is a compactly supported and  $S+1$  times continuously differentiable function, then

$$\|\sigma_n(\alpha, \beta; h, f)\|_p \leq c \|f\|_p. \quad (5.5)$$

Using the same methods as in the trigonometric case and the above results, one can easily obtain direct and converse theorems as well as the wavelet-like representation theorems for characterization of suitably defined global smoothness classes; these results are formulated in [59]. However, these bounds by themselves are not sufficient to obtain a characterization of local smoothness. We proved the following localization estimates on the kernels  $\Phi_n$ , see [45, 48].

**Theorem 5.2** Let  $\alpha, \beta \geq -1/2$ ,  $S \in \mathbb{N}$ , and let  $h_j = 0$  for all sufficiently large  $j$ . Then

$$a \left| \sum_{k=0}^{\infty} h_j p_j^{(\alpha, \beta)}(\cos \theta) p_j^{(\alpha, \beta)}(\cos \varphi) \right| \leq c \sum_{j=0}^{\infty} \min \left( (j+1), \frac{1}{|\theta - \varphi|} \right)^{\max(\alpha, \beta) + S + 1/2} \times \sum_{m=1}^S (j+1)^{\max(\alpha, \beta) + 1/2 - S + m} |\Delta^m h_j|, \quad (5.6)$$

for  $\theta, \varphi \in [0, \pi]$ . In particular, if  $h : [0, \infty) \rightarrow [0, \infty)$  is a compactly supported function that can be expressed as an  $S$ -times iterated integral of a function of bounded total variation  $V$ , and if  $h'(t) = 0$  in a neighborhood of 0, then

$$|\Phi_n(\alpha, \beta; h, \cos \theta, \cos \varphi)| \leq c n^{2\max(\alpha, \beta) + 2} V \min \left( 1, \frac{1}{(n|\theta - \varphi|)^{\max(\alpha, \beta) + S + 1/2}} \right), \quad (5.7)$$

for  $n \in \mathbb{N}$ .

A wavelet-like representation with characterization of local smoothness classes is given in [48].

The subject of Marcinkiewicz–Zygmund inequalities in this context is very well studied. Perhaps, the most classical example is a simple consequence of the Gauss–Jacobi quadrature formula. For  $n \in \mathbb{N}$ , let  $\{x_{k,n}\}_{k=1}^n$  be the zeros of  $P_n^{(\alpha, \beta)}$ , and let

$$\lambda_{k,n} \stackrel{\text{def}}{=} \left\{ \sum_{j=0}^{n-1} p_j^{(\alpha, \beta)}(x_{k,n})^2 \right\}^{-1}, \quad k = 1, 2, \dots, n.$$

Then the Gauss–Jacobi quadrature formula implies that

$$\sum_{k=1}^n \lambda_{k,n} |P(x_{k,n})|^2 = \int_{-1}^1 |P(y)|^2 w_{\alpha, \beta}(y) dy, \quad P \in \Pi_{n-1}. \quad (5.8)$$

The following analogue in the case of  $L^p$  norms was proved in [68, Theorem 25, p. 168].

**Theorem 5.3** Let  $1 \leq p < \infty$  and let  $c_1 > 0$ . Then there exists a constant  $c$  depending only on  $\alpha, \beta$ , and  $c_1$ , such that for all  $m \in \mathbb{N}$  and  $n \in \mathbb{N}$  with  $1 \leq m \leq c_1 n$ ,

$$\sum_{k=1}^n \lambda_{k,n} |P(x_{k,n})|^p \leq c \|P\|_{\alpha, \beta; p}^p, \quad P \in \Pi_m. \quad (5.9)$$

This theorem found deep applications in investigations related to weighted mean convergence of Lagrange interpolation [67, 68]. A survey of many of the classical results in this direction and their applications can be found in the paper [38] by Lubinsky. In [48], we proved the existence of M–Z quadrature measures based on arbitrary set of points on  $[-1, 1]$  subject to a density condition, and we gave applications to wavelet-like representations based on values of the function at these points.

Characteristically for polynomial approximation, one can also construct localized operators which yield approximation commensurate with analyticity of the target function on intervals, rather than the much weaker smoothness conditions studied in the previous section. If  $1 \leq p \leq \infty$ ,  $n \geq 0$ , and  $f \in L^p(\alpha, \beta)$ , then we define the degree of best weighted approximation of  $f$  by polynomials of degree at most  $n$  by

$$E_{n,p}(\alpha, \beta; f) \stackrel{\text{def}}{=} \min_{P \in \Pi_n} \|f - P\|_{\alpha, \beta; p}.$$

For integer  $n \geq 1$ , let the numbers  $H_{j,n}^*$ ,  $j = 0, \dots, 5n-1$  be defined by

$$\left(\frac{1+x}{2}\right)^n \Phi_{4n}(\alpha, \beta; x, 1) = \sum_{j=0}^{5n-1} H_{j,n}^* p_j^{(\alpha, \beta)}(x) p_j^{(\alpha, \beta)}(1), \quad (5.10)$$

and let

$$\sigma_n(\alpha, \beta; H^*, f, x) \stackrel{\text{def}}{=} \sum_{j=0}^{5n-1} H_{j,n}^* \hat{f}(j) p_j^{(\alpha, \beta)}(x), \quad x \in [-1, 1].$$

In [19, Theorem 3.3], we proved the following.

**Theorem 5.4** (a) *Let  $1 \leq p \leq \infty$ ,  $\alpha, \beta \geq -1/2$ , and let  $f \in L^p(\alpha, \beta)$ . Then, with  $H^*$  as defined in (5.10), we have  $\sigma_n(\alpha, \beta; H^*, P) = P$  for  $P \in \Pi_n$ , and  $\sup_{n \in \mathbb{N}} \|\sigma_n(\alpha, \beta; H^*, f)\|_{\alpha, \beta; p} \leq c \|f\|_{\alpha, \beta; p}$ . In addition,*

$$E_{5n,p}(\alpha, \beta; f) \leq \|f - \sigma_n(\alpha, \beta; H^*, f)\|_{\alpha, \beta; p} \leq c_1 E_{n,p}(\alpha, \beta; f). \quad (5.11)$$

(b) *Let  $f \in C([-1, 1])$ ,  $x_0 \in [-1, 1]$ , and let  $f$  have an analytic continuation to a complex neighborhood of  $x_0$ , given by  $\{z \in \mathbb{C} : |z - x_0| \leq d\}$  for some  $d$  with  $0 < d \leq 2$ . Then*

$$|f(x) - \sigma_n(\alpha, \beta; H^*, f, x)| \leq c(f, x_0) \exp\left(-c_1(d)n \frac{d^2 \log(e/2)}{e^2 \log(e^2/d)}\right),$$

$$x \in [x_0 - d/e, x_0 + d/e] \cap [-1, 1], \quad (5.12)$$

where  $\log$  is the natural logarithm, and  $e$  is the basis of this logarithm.

## 5.2 Approximation on the sphere

Theorems about Jacobi expansions translate easily into analogous theorems for approximation on the sphere. The following paragraph is taken from [61]. Let  $q \in \mathbb{N}$ , and let

$$\mathbb{S}^q \stackrel{\text{def}}{=} \left\{ (x_1, \dots, x_{q+1}) \in \mathbb{R}^{q+1} : \sum_{j=1}^{q+1} x_j^2 = 1 \right\}.$$

A spherical cap, centered at  $\mathbf{x}_0 \in \mathbb{S}^q$  and with radius  $\alpha$  is defined by

$$\mathbb{S}_\alpha^q(\mathbf{x}_0) \stackrel{\text{def}}{=} \{ \mathbf{x} \in \mathbb{S}^q : \mathbf{x} \cdot \mathbf{x}_0 \geq \cos \alpha \}.$$

We note that for all  $\mathbf{x}_0 \in \mathbb{S}^q$  we have  $\mathbb{S}_\pi^q(\mathbf{x}_0) = \mathbb{S}^q$ . In this subsection, the surface area (aka volume element) measure on  $\mathbb{S}^q$  will be denoted by  $\mu_q^*$  (there being no chance of confusion with the notation on the torus), and we write  $\omega_q \stackrel{\text{def}}{=} \mu_q^*(\mathbb{S}^q)$ . The spaces  $X^p(\mathbb{S}^q)$  and  $C(\mathbb{S}^q)$  on the sphere are defined analogously to the case of the interval.

A spherical polynomial of degree  $m$  is the restriction to  $\mathbb{S}^q$  of a polynomial in  $q+1$  real variables with total degree  $m$ . For integer  $n \geq 0$ , the class of all spherical polynomials of degree at most  $n$  will be denoted by  $\Pi_n^q$ . As before, we extend this notation for non-integer values of  $n$  by setting  $\Pi_n^q \stackrel{\text{def}}{=} \Pi_{[n]}^q$ . For integer  $\ell \geq 0$ , the class of all homogeneous, harmonic, spherical polynomials of degree  $\ell$  will be denoted by  $\mathbf{H}_\ell^q$ , and its dimension by  $d_\ell^q$ . For each integer  $\ell \geq 0$ , let  $\{Y_{\ell,k} : k = 1, 2, \dots, d_\ell^q\}$  be a  $\mu_q^*$ -orthonormalized basis for  $\mathbf{H}_\ell^q$ . It is known that for any integer  $n \geq 0$ , the system  $\{Y_{\ell,k} : \ell = 0, 1, \dots, n, k = 1, 2, \dots, d_\ell^q\}$  is an orthonormal basis for  $\Pi_n^q$ , cf. [63, 75]. The connection with the theory of orthogonal polynomials on  $[-1, 1]$  is the following addition formula

$$\sum_{k=1}^{d_\ell^q} Y_{\ell,k}(\mathbf{x}) Y_{\ell,k}(\mathbf{y}) = \omega_{q-1}^{-1} p_\ell^{(q/2-1, q/2-1)}(1) p_\ell^{(q/2-1, q/2-1)}(\mathbf{x} \cdot \mathbf{y}), \quad \ell = 0, 1, \dots,$$

cf. [63] where the notation is different.

Analogues of the direct and converse theorems in the case of approximation on the sphere are given, for instance, by Pawelke [71] and Lizorkin and Rustamov [35]. The existence of M–Z quadrature measures was proved in [54]. Numerical constructions and various experiments to demonstrate the effectiveness of the localized operators are given in [34]. A wavelet-like representation including local smoothness classes is given in [47]. In the case of spherical caps, the existence of M–Z quadrature measures is given in [46, 9]; the corresponding results on spherical triangles are proved in [44] and numerical constructions are given in [2].

The analogue of the PBF network in this context is the so-called *zonal function (ZF) network*. A zonal function network is a function of the form  $\mathbf{x} \mapsto \sum_{k=1}^n c_k \phi(\mathbf{x} \cdot \mathbf{y}_k)$ , where  $\mathbf{x}$  and the  $\mathbf{y}_k$ 's are on  $\mathbb{S}^q$  and  $\phi \in L^1(q/2 - 1, q/2 - 1)$ . We observe that

analogous to the “Mercer expansion”

$$G(\mathbf{x} - \mathbf{y}) = \sum_{\mathbf{k}} \hat{G}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x}) \overline{\exp(i\mathbf{k} \cdot \mathbf{y})},$$

for the activation function  $G$  of a PBF network, one has the expansion

$$\phi(\mathbf{x} \cdot \mathbf{y}) = \sum_{\ell=0}^{\infty} \hat{\phi}(\ell) \sum_{k=1}^{d_{\ell}^q} Y_{\ell,k}(\mathbf{x}) Y_{\ell,k}(\mathbf{y})$$

for the activation function  $\phi$  of a ZF network. The ideas in Sect. 4 can be carried over almost verbatim to the case of ZF networks. In particular, the direct and converse theorems in this connection are obtained in [55,56]. The wavelet-like representation for ZF network frames was announced by Shvarts in a joint paper with HNM in a meeting in Barcelona, Spain, in December, 2011.

**Acknowledgments** We thank the referee for his comments, corrections, critical observations, valuable suggestions, and, in particular, for pointing out [14,32,40,41,72].

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

1. Askey, R., Wainger, S.: A convolution structure for Jacobi series. *Am. J. Math.* **91**(2), 463–485 (1969)
2. Beckman, J., Mhaskar, H.N., Prestin, J.: Quadrature formulas for integration of multivariate trigonometric polynomials on spherical triangles. *Int. J. Geomath.* (accepted)
3. Blackmore, K.L., Williamson, R.C., Mareels, I.M.Y.: Decision region approximation by polynomials or neural networks. *IEEE Trans. Inf. Theory* **43**(3), 903–907 (1997)
4. Carleson, L.: On convergence and growth of partial sums of Fourier series. *Acta Math.* **116**, 135–157 (1966)
5. Chandrasekharan, S., Jayaraman, K., Mhaskar, H.N.: Minimum Sobolev norm interpolation with trigonometric polynomials on the torus. *J. Comput. Phys.* (accepted)
6. Cheney, E.W.: *Introduction to Approximation Theory*, 2nd edn. Chelsea Publishing Company, New York (1982)
7. Chui, C.K., Mhaskar, H.N.: Smooth function extension based on high dimensional unstructured data. *Math. Comput.* (accepted)
8. Czipser, J., Freud, G.: Sur l’approximation d’une fonction périodique et ses dérivées successives par un polynôme trigonométrique et par ses dérivées successives. *Acta Math.* **5**, 285–290 (1957)
9. Dai, F., Wang, H.: Positive cubature formulas and Marcinkiewicz–Zygmund inequalities on spherical caps. *Constr. Approx.* **31**(1), 1–36 (2010)
10. Daubechies, I.: *Ten Lectures on Wavelets*, CBMS-NSF Series in Appl. Math., SIAM Publ., Philadelphia (1992)
11. DeVore, R., Howard, R., Micchelli, C.A.: Optimal nonlinear approximation. *Manuscr. Math.* **63**, 469–478 (1989)
12. DeVore, R.A., Lorentz, G.G.: *Constructive Approximation*. Springer, Berlin (1993)
13. Dick, J., Pillichshammer, F.: *Digital Nets and Sequences*. Cambridge University Press, Cambridge (2010)
14. Dung, D., Micchelli, C.A.: Sparse grid approximation by translates of the Korobov function (arXiv:1212.6160 [math.FA])

15. Dyn, N.: Interpolation and approximation by radial and related functions. In: Chui, K., Schumaker, L.L., Ward, J.D. (eds.) *Approximation Theory VI*, vol. 1, pp. 211–234. Academic Press, Boston (1989)
16. Erb, W.: Optimally space localized polynomials with applications in signal processing. *J. Fourier Anal. Appl.* **18**(1), 45–66 (2012)
17. Erdélyi, A., Magnus, W., Oberhettinger, F., Tricomi, F.G.: *Higher Transcendental Functions*, vol. 1. McGraw Hill, New York (1953)
18. Fejér, L.: Untersuchungen über Fouriersche Reinen. *Math. Ann.* **58**, 51–69 (1904)
19. Filbir, F., Mhaskar, H.N., Prestin, J.: On a filter for exponentially localized kernels based on Jacobi polynomials. *J. Approx. Theory* **160**, 256–280 (2009)
20. Filbir, F., Mhaskar, H.N.: A quadrature formula for diffusion polynomials corresponding to a generalized heat kernel. *J. Fourier Anal. Appl.* **16**, 629–657 (2010)
21. Filbir, F., Mhaskar, H.N.: Marcinkiewicz–Zygmund measures on manifolds. *J. Complex.* **27**, 568–596 (2011)
22. Filbir, F., Mhaskar, H.N., Prestin, J.: On the problem of parameter estimation in exponential sums. *Constr. Approx.* **35**(3), 323–343 (2012)
23. Fröhlich, J., Uhlmann, M.: Orthonormal polynomial wavelets on the interval and applications to the analysis of turbulent flow fields. *Siam J. Appl. Math.* **63**, 1789–1830 (2003)
24. Girgensohn, R., Prestin, J.: Lebesgue constants for an orthogonal polynomial Schauder basis. *J. Comput. Anal. Appl.* **2**, 159–175 (2000)
25. Girosi, F., Jones, M., Poggio, T.: Regularization theory and neural networks architectures. *Neural Comput.* **7**, 219–269 (1995)
26. Hardy, G.H.: *Divergent Series*, AMS Chelsea Publ., American Mathematical Society, Providence (1991)
27. Hardy, R.L.: Theory and applications of the multiquadric-biharmonic method, 20 years of discovery 1968–1988. *Comput. Math. Appl.* **19**(8–9), 163–208 (1990)
28. Hassanieh, H., Indyk, P., Katabi, D., Price, E.: Simple and Practical Algorithm for Sparse Fourier Transform, ACM-SIAM Symposium on discrete algorithms, SIAM publ., pp. 1183–1194 (2012)
29. Horn, R.A., Johnson, C.R.: *Matrix Analysis*. Cambridge University Press, Cambridge (1985)
30. Hunt, R.: On the convergence of Fourier series. *Proc. Conf. Edwardsville*, pp. 235–255 (1967)
31. Kilgore, T.J., Prestin, J., Selig, K.: Orthogonal algebraic polynomial Schauder bases of optimal degree. *J. Fourier Anal. Appl.* **2**, 597–610 (1996)
32. Kolmogorov, A.N.: Über die beste Annäherung von Funktionen einer Funktionklasse. *Ann. Math.* **37**, 107–111 (1936)
33. Lavretsky, E., Hovakimyan, N., Calise, A.: Adaptive Extremum Seeking Control Design. American Control Conference (2003)
34. Le Gia, Q.T., Mhaskar, H.N.: Localized linear polynomial operators and quadrature formulas on the sphere. *SIAM J. Numer. Anal.* **47**(1), 440–466 (2008)
35. Lizorkin, P.I., Rustamov, Kh.P.: Nikolskii–Besov spaces on the sphere in connection with approximation theory. *Tr. Mat. Inst. Steklova* **204**, 172–201 (1993) (*Proc. Steklov Inst. Math.* **3**, 149–172 (1994))
36. Lorentz, G.G., Golitschek, M.v., Makovoz, Y.: *Constructive Approximation*, Advanced Problems. Springer, Berlin (1996)
37. Lorentz, R.A., Sahakian, A.A.: Orthogonal trigonometric Schauder bases of optimal degree for  $C(0, 2\pi)$ . *J. Fourier Anal. Appl.* **1**, 103–112 (1994)
38. Lubinsky, D.S.: Marcinkiewicz inequalities: methods and results. In: Milovanovic, G.V. (ed.) *Recent Progress in Inequalities*, pp. 213–240. Kluwer, Dordrecht (1998)
39. Maggioni, M., Mhaskar, H.N.: Diffusion polynomial frames on metric measure spaces. *Appl. Comput. Harm. Anal.* **24**(3), 329–353 (2008)
40. Maiorov, V.: Almost optimal estimates for best approximation by translates on a torus. *Constr. Approx.* **21**, 1–20 (2005)
41. Mathé, P.:  $s$ -Number in information-based complexity. *J. Complex.* **6**, 41–66 (1990)
42. Mhaskar, H.N.: When is approximation by Gaussian networks necessarily a linear process? *Neural Netw.* **17**, 989–1001 (2004)
43. Mhaskar, H.N.: Approximation theory and neural networks. In: Jain, K., Krishnan, M., Mhaskar, H.N., Prestin, J., Singh, D. (eds.) *Wavelet Analysis and Applications*, Proceedings of the international workshop in Delhi, 1999, pp. 247–289. Narosa Publishing, New Delhi (2001)

44. Mhaskar, H.N.: Local quadrature formulas on the sphere, II. In: Neamtu, M., Saff, E.B. (eds.) *Advances in Constructive Approximation*, pp. 333–344. Nashboro Press, Nashville (2004)
45. Mhaskar, H.N.: Polynomial operators and local smoothness classes on the unit interval. *J. Approx. Theory* **131**, 243–267 (2004)
46. Mhaskar, H.N.: Local quadrature formulas on the sphere. *J. Complex.* **20**, 753–772 (2004)
47. Mhaskar, H.N.: On the representation of smooth functions on the sphere using finitely many bits. *Appl. Comput. Harmon. Anal.* **18**(3), 215–233 (2005)
48. Mhaskar, H.N.: Polynomial operators and local smoothness classes on the unit interval, II. *Jaén J. Approx.* **1**(1), 1–25 (2005)
49. Mhaskar, H.N.: Eignets for function approximation. *Appl. Comput. Harmon. Anal.* **29**, 63–87 (2010)
50. Mhaskar, H.N., Micchelli, C.A.: Approximation by superposition of a sigmoidal function and radial basis functions. *Adv. Appl. Math.* **13**, 350–373 (1992)
51. Mhaskar, H.N., Micchelli, C.A.: Degree of approximation by neural and translation networks with a single hidden layer. *Adv. Appl. Math.* **16**, 151–183 (1995)
52. Mhaskar, H.N., Micchelli, C.A.: Dimension-independent bounds on approximation by neural networks. *IBM J. Res. Dev.* **38**, 277–284 (1994)
53. Mhaskar, H.N., Micchelli, C.A.: How to choose an activation function. In: Cowan, J.D., Tesauro, G., Alspector, J. (eds.) *Neural Information Processing Systems*, vol. 6, pp. 319–326. Morgan Kaufmann Publishers, San Francisco (1993)
54. Mhaskar, H.N., Narcowich, F.J., Ward, J.D.: Spherical Marcinkiewicz–Zygmund inequalities and positive quadrature. *Math. Comput.* **70**(235), 1113–1130 (2001) (Corrigendum. *Math. Comput.* **71**, 453–454 (2001))
55. Mhaskar, H.N., Narcowich, F.J., Ward, J.D.: Approximation properties of zonal function networks using scattered data on the sphere. *Adv. Comput. Math.* **11**, 121–137 (1999)
56. Mhaskar, H.N., Narcowich, F.J., Prestin, J., Ward, J.D.:  $L^p$  Bernstein estimates and approximation by spherical basis functions. *Math. Comput.* **79**(271), 1647–1679 (2010)
57. Mhaskar, H.N., Pai, D.V.: *Fundamentals of Approximation Theory*. Alpha Science Intl., Revised Edition (2007)
58. Mhaskar, H.N., Prestin, J.: On the detection of singularities of a periodic function. *Adv. Comput. Math.* **12**, 95–131 (2000)
59. Mhaskar, H.N., Prestin, J.: Polynomial frames: a fast tour. In: Chui, K., Neamtu, M., Schumaker, L. (eds.) *Approximation Theory XI: Gatlinburg 2004*, *Mod. Methods Math*, pp. 287–318. Nashboro Press, Brentwood (2005)
60. Mhaskar, H.N., Prestin, J.: On local smoothness classes of periodic functions. *J. Fourier Anal. Appl.* **11**(3), 353–373 (2005)
61. Mhaskar, H.N., Prestin, J.: Polynomial operators for spectral approximation of piecewise analytic functions. *Appl. Comput. Harmon. Anal.* **26**, 121–142 (2009)
62. Micchelli, C.A.: Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constr. Approx.* **2**(1), 11–22 (1986)
63. Müller, C.: *Spherical Harmonics*, *Lecture Notes in Mathematics*, vol. 17. Springer, Berlin (1966)
64. Narcowich, F.J., Ward, J.D.: Norms of inverses and condition numbers for matrices associated with scattered data. *J. Approx. Theory* **64**, 69–94 (1991)
65. Narcowich, F.J., Ward, J.D.: Norm estimates for inverses of a general class of scattered data radial function interpolation matrices. *J. Approx. Theory* **69**, 84–109 (1992)
66. Natanson, I.P.: *Constructive Function Theory*, vol. I. Frederick Ungar, New York (1964)
67. Nevai, P.: Mean convergence of Lagrange interpolation. *J. Approx. Theory* **18**, 363–377 (1976)
68. Nevai, P.: *Orthogonal Polynomials*. *Mem. Amer. Math. Soc.*, vol. 203. American Mathematical Society, Providence (1979)
69. Nevai, P.: The incredible but, I swear, true story of  $n$  vs.  $2n$  in the Bernstein Inequality (2013, in preparation)
70. Park, J., Sandberg, I.W.: Universal approximation using radial basis function networks. *Neural Comput.* **3**, 246–257 (1991)
71. Pawelke, S.: Über die Approximationsordnung bei Kugelfunktionen und algebraischen Polynomen. *Tôhoku Math. J.* **24**, 473–486 (1972)
72. Pinkus, A.:  *$n$ -Widths in Approximation Theory*. Springer, Berlin (1985)
73. Schaback, R., Wendland, H.: Inverse and saturation theorems for radial basis function interpolation. *Math. Comput.* **71**, 669–681 (2002)



74. Stechkin, S.B.: The approximation of periodic functions by Fejr sums. (Russian). Trudy Mat. Inst. Steklov. **62**, 48–60 (1961)
75. Stein, E.M., Weiss, G.: Fourier Analysis on Euclidean Spaces. Princeton University Press, Princeton (1971)
76. Szegő, G.: Orthogonal Polynomials, Amer. Math. Soc. Colloq. Publ., vol. 23. American Mathematical Society, Providence (1975)
77. Timan, A.F.: Theory of Approximation of Functions of a Real Variable. Pergamon Press, Oxford (1963; English translation)
78. Zeevi, A., Meir, R., Maiorov, V.: Approximation and estimation bounds for nonlinear regression using mixtures of experts. Neural Netw. **10**(1), 99–109 (1997)
79. Zygmund, A.: Trigonometric Series. Cambridge University Press, Cambridge (1977)